Maintaining distance within the genome : the role of Sen1 and RNases H

Umberto Aiello

Thèse de Doctorat de Université de Paris

École doctorale **562 Bio Sorbonne Paris Cité**

Préparée à l'**Institut Jacques Monod**

Maintaining distance within the genome: the role of Sen1 and RNases H

Par **Umberto Aiello**

Thèse de Doctorat en **Génétique**

dirigée par **Domenico Libri**

Présentée et soutenue publiquement à l'Institut Jacques Monod, le 01 Décembre 2021

Devant un jury composé de :

Président du jury

Jonathan Weitzman, Professeur, Université de Paris

Directeur de thèse

Domenico Libri, Directeur de Recherche, CNRS

Rapporteurs

Sarah Lambert, Directeur de Recherche, Institut Curie

Vincent Vanoosthuyse, Directeur de Recherche, ENS Lyon

Examinateurs

Michelle Debatisse, Professeur Émérite, Sorbonne Université

ACKNOWLEDGMENTS

Doing a PhD offers many advantages, but nothing good comes without challenges. Many difficulties, indeed, that I would have not been able to overcome alone. First, all this would have not been possible without the unlimited and unconditional support of my family, and it is to you that I want to dedicate this thesis. You have believed in me and offered all what you could give.

I want to express all my gratitude to Domenico, who has welcomed me in his lab and allowed me to work on this project. I am so glad that you have been my thesis director and mentor. I believe that I have learned so much from you, and not only on how to be a better scientist, but also on how to be a better person.

I want to thank all past and current members of the Libri lab who have created such a stimulating environment. I could always ask for help to any of you and never find a closed door and the scientific discussion that you have provided has been a ground where I could exercise my mind and grow as a reasoner.

I also would like to acknowledge the members of the jury, Michelle Debatisse, Sarah Lambert, Vincent Vanoosthuyse and Jonathan Weitzman, for having accepted to evaluate my work and dedicate some of their time to me. This was deeply appreciated.

I would like to thank Armelle Lengonne, Philippe Pasero and the Palancade lab for the stimulating discussions, and particularly Benoît, whose help was an important contribution for the success of this thesis project.

Last but not least, all my friends, inside and outside the Institute: thank you so much. Nothing is possible without laughs and beers, especially a PhD!

ABSTRACT

Intrinsic damage potential from transcription originates from different sources, among which Transcription-Replication Conflicts (TRCs). The transcription process is tightly regulated, both at the level of initiation and termination to avoid risks of conflict with other DNA-associated processes. Nevertheless, many transcription events occur in the genome that are seemingly non-functional, which defines the concept of pervasive transcription and underscores the challenging necessity for the cell to coordinate transcription with other processes.

TRCs take place upon encountering of the transcription and the replication machineries, and can lead to replication fork stalling and potentially DNA damage. TRCs are linked to the formation of R-loops, which are intermediate heteroduplexes formed when the nascent RNA hybridizes to its own DNA template in the wake of the polymerase, thus extruding a single stranded DNA molecule. R-loops have both physiological functions as intermediates of some cellular processes, and a genotoxic effect when their formation is unscheduled. Moreover, they are an impediment to the progression of replicative forks, being able by themselves to induce fork stalling.

The evolutionary conserved helicase Sen1 (SETX in human) plays a prominent role in limiting and resolving TRCs. On the one hand, as component of the Nrd1-Nab3-Sen1 (NNS) complex, Sen1 is an essential transcription termination factor, responsible for dislodging the transcriptional apparatus from thousands of non-coding genes (i.e., Cryptic Unstable Transcripts or CUTs). On the other hand, Sen1 has been proposed to prevent R-loops accumulation by virtue of its DNA:RNA unwinding activity. Currently, the specific mechanisms by which Sen1 affects R-loops levels and/or TRCs is unknown, and previous works have been performed with *sen1* mutants that strongly also affect its transcription termination function.

Here, we focused on the dissection of the role of Sen1 in genome stability, aiming at separating the functions of Sen1 in non-coding RNA genes termination and at R-loops and/or TRCs. In the frame of a collaborative work, we have shown that Sen1 interacts directly with the replisome and participated in characterizing a mutant (*sen1-3*) that harbours three point mutations that completely abolish binding of Sen1 to the replisome, yet leaving untouched its ability to terminate transcription via the NNS pathway. Importantly, RNase H activity,

brought by either one of the two yeast RNase H enzymes, is essential in *sen1-3* cells, thus suggesting that Sen1 plays a role at the replisome that might be redundant or concurrent with the one of RNases H. These are ribonucleases that have been shown to have a critical function in R-loops clearance, thanks to their ability to recognize and digest DNA:RNA hybrids.

By a systematic and comprehensive genomic approach, we investigated the role of Sen1 and its link with RNases H, unveiling that Sen1 promotes removal of RNA polymerase II at many locations where conflicts with replication or other processes occur. We have also studied the role of RNases H in these processes and the impact of R-loops. To this end we have developed a novel method, H-CRAC, to detect RNase H targets, possibly R-loops, with nucleotide resolution. We show that H-CRAC signals significantly overlap available lower resolution R-loops maps and provide evidence that at least a fraction of R-loops can be detected by this method. We propose a model that integrates the roles of RNases H and Sen1 in the resolution of replication- and transcription-transcription conflicts.

Keywords: transcription-replication conflicts; R-loops; Sen1; RNase H; genome instability.

RESUMÉ

Le potentiel de dommages liés intrinsèquement à la transcription provient de différentes sources, parmi lesquelles les conflits entre transcription et réplication (CTR). Le processus de transcription est étroitement régulé, tant au niveau de l'initiation que de la terminaison, afin d'éviter les risques associés aux conflits avec d'autres processus liés à l'ADN. Néanmoins, de nombreux événements de transcription se produisent dans le génome alors qu'ils sont apparemment non fonctionnels, ce qui définit le concept de transcription omniprésente et souligne la nécessité pour la cellule de coordonner la transcription avec d'autres processus.

Les CTR se produisent lors de la rencontre entre les machineries de transcription et de réplication, entraînant un blocage de la fourche de réplication et provoquant potentiellement des dommages à l'ADN. Les CTR sont liés à la formation de R-loops, qui sont des hétéroduplexes intermédiaires formés lorsque l'ARN naissant s'hybride à sa propre matrice d'ADN, extrudant ainsi une molécule d'ADN simple brin. Les R-loops ont à la fois des fonctions physiologiques en tant qu'intermédiaires de certains processus cellulaires et un effet génotoxique lorsque leur formation n'est pas programmée. De plus, ils sont un obstacle à la progression des fourches réplicatives, étant capables par eux-mêmes d'induire un blocage de la fourche.

L'hélicase Sen1 (SETX chez l'homme), conservée au cours de l'évolution, joue un rôle important dans la limitation et la résolution des CTR. D'une part, en tant que composant du complexe Nrd1-Nab3-Sen1 (NNS), Sen1 est un facteur essentiel de terminaison de la transcription, fonctionnant sur de milliers de gènes non codants (c'est-à-dire des transcrits cryptiques instables ou CUTs). D'autre part, il a été proposé que Sen1 empêche l'accumulation de R-loops en vertu de son activité de déroulement de l'ADN:ARN. Actuellement, les mécanismes spécifiques par lesquels Sen1 affecte les niveaux de R-loops et des CTR ne sont pas bien définis, et la plupart des expériences précédentes ont été réalisées avec des mutants de Sen1 qui affectent fortement aussi sa fonction de terminaison de la transcription.

Notre travail porte sur la dissection du rôle de Sen1 dans la stabilité du génome, visant à séparer les fonctions de Sen1 dans la terminaison des gènes d'ARN non codant et au niveau des R-loops et des CTR. Dans le cadre d'un travail collaboratif, nous avons montré que Sen1 interagit directement avec le replisome et nous avons caractérisé un mutant (sen1-3) qui porte trois mutations ponctuelles qui abolissent la liaison de Sen1 au replisome, tout en laissant

intacte sa capacité à terminer la transcription via la voie NNS. Il est important de noter que l'activité de RNase H, apportée par l'une ou l'autre des deux enzymes RNase H de levure, est essentielle dans les cellules sen1-3, ce qui suggère que Sen1 joue un rôle au niveau du replisome qui pourrait être redondant ou concurrent à celui des RNases H. Ces facteurs sont des ribonucléases qui ont une fonction critique dans l'élimination des R-loops, grâce à leur capacité à reconnaître et à digérer les hybrides ADN:ARN.

Par une approche génomique systématique, nous avons étudié le rôle de Sen1 et son lien avec les RNases H, révélant que Sen1 favorise l'élimination de l'ARN polymérase II à de nombreux endroits où se produisent des conflits avec la réplication ou d'autres processus. Nous avons également étudié le rôle des RNases H dans ces processus et l'impact des R-loops. A cette fin, nous avons développé une nouvelle méthode, H-CRAC, pour détecter les cibles RNase H avec une résolution nucléotidique. Nous montrons que les signaux H-CRAC chevauchent de manière significative les cartes R-loop connues de résolution plus faible et nous apportons la preuve qu'au moins une fraction des R-loops peut être détectée par cette méthode. Nous proposons un modèle qui intègre les rôles des RNases H et Sen1 dans la résolution des CTR et sur d'autres sites de conflits.

Mots-clés : conflits entre transcription et réplication ; R-loops ; Sen1 ; RNase H ; instabilité du genome.

RESUMÉ DÉTAILLÉ

Le potentiel de dommages liés intrinsèquement à la transcription provient de différentes sources, parmi lesquelles les conflits entre transcription et réplication (CTR). Le processus de transcription est étroitement régulé, tant au niveau de l'initiation que de la terminaison, afin d'éviter les risques associés aux conflits avec d'autres processus liés à l'ADN. Néanmoins, de nombreux événements de transcription se produisent dans le génome alors qu'ils sont apparemment non fonctionnels, ce qui définit le concept de transcription omniprésente et souligne la nécessité pour la cellule de coordonner la transcription avec d'autres processus.

Depuis la découverte que la transcription et la réplication ne sont pas séparées dans le temps ou l'espace, la façon dont les cellules coordonnent ces deux processus est une question importante. Les CTR se produisent lors de la rencontre entre les machineries de transcription et de réplication, entraînant un blocage de la fourche de réplication et provoquant potentiellement des dommages à l'ADN. Les CTR sont liés à la formation de R-loops, qui sont des hétéroduplexes intermédiaires formés lorsque l'ARN naissant s'hybride à sa propre matrice d'ADN, extrudant ainsi une molécule d'ADN simple brin. Les R-loops ont à la fois des fonctions physiologiques en tant qu'intermédiaires de certains processus cellulaires, notamment la commutation isotypique ou la réplication du génome mitochondrial, et un effet génotoxique lorsque leur formation n'est pas programmée. De plus, ils sont un obstacle à la progression des fourches réplicatives, étant capables par eux-mêmes d'induire un blocage de la fourche. Au vu de la genotoxicité potentielle de ces structures, la formation des R-loops est étroitement contrôlée au niveau cellulaire, et plusieurs mécanismes sont en place pour prévenir leur formation et si pour les dégrader. Par exemple, leur formation est elle-même limitée par les différentes étapes qui suivent la synthèse de l'ARN et qui permettent sa maturation, son assemblage en ribonucléoparticule et son export hors du noyau.

L'hélicase Sen1 (SETX chez l'homme), conservée au cours de l'évolution, joue un rôle important dans la limitation et la résolution des CTR. D'une part, en tant que composant du complexe Nrd1-Nab3-Sen1 (NNS), Sen1 est un facteur essentiel de terminaison de la transcription, fonctionnant sur de milliers de gènes non codants (c'est-à-dire des transcrits cryptiques instables ou CUTs). D'autre part, il a été proposé que Sen1 empêche l'accumulation de R-loops en vertu de son activité de déroulement de l'ADN:ARN. En effet, il a été démontré que l'élimination de Sen1 entraîne une instabilité génomique, sous-tendue par une augmentation de la recombinaison associée à la transcription, des niveaux des R-loops et une

stabilité de fourche affectée. Malgré tout, actuellement les mécanismes spécifiques par lesquels Sen1 affecte les niveaux de R-loops et des CTR ne sont pas bien définis, et la plupart des expériences précédentes ont été réalisées avec des mutants de Sen1 qui affectent fortement aussi sa fonction de terminaison de la transcription, ce qui peut entraîner plusieurs effets indirects.

Notre travail porte sur la dissection du rôle de Sen1 dans la stabilité du génome, visant à séparer les fonctions de Sen1 dans la terminaison des gènes d'ARN non codant et au niveau des R-loops et des CTR. En effet, nous montrons que la déplétion de Sen1 entraîne des changements majeurs dans le transcriptome, avec l'activation de plusieurs gènes de la réponse au stress. De plus, l'occupation de l'ARN polymérase II dans les régions intergéniques non transcrites est fortement augmentée, favorisant ainsi les possibilités de former des R-loops ou d'avoir des CTR.

Dans le cadre d'un travail collaboratif, nous avons montré que Sen1 interagit directement avec le réplisome via son domaine N-terminal et nous avons caractérisé un mutant (*sen1-3*) qui porte trois mutations ponctuelles qui abolissent la liaison de Sen1 au réplisome, tout en laissant intacte sa capacité à terminer la transcription via la voie NNS. Il est important de noter que l'activité de RNase H, apportée par l'une ou l'autre des deux enzymes RNase H de levure, est essentielle dans les cellules *sen1-3*, ce qui suggère que Sen1 joue un rôle au niveau du réplisome qui pourrait être redondant ou complémentaire à celui des RNases H. Ces facteurs sont des ribonucléases qui ont une fonction critique dans l'élimination des R-loops, grâce à leur capacité à reconnaître et à digérer les hybrides ADN:ARN. Toutefois, et de façon importante, les cellules *sen1-3* ne présentent pas d'augmentation de la recombinaison associée à la transcription ou des niveaux de R-loops, ce qui suggère que les phénotypes d'instabilité génomique associés à la déplétion de Sen1 pourraient être dus à un défaut simultané dans la terminaison de la transcription et à un rôle direct dans la stabilité du génome.

Par une approche génomique systématique, nous avons étudié le rôle de Sen1, en utilisant les avantages inédits offerts par l'allèle de séparation de fonction *sen1-3*, et son lien avec les RNases H, révélant que Sen1 favorise l'élimination de l'ARN polymérase II à de nombreux endroits où se produisent des conflits avec la réplication ou d'autres processus. Nous avons observé que l'absence de Sen1 au niveau du réplisome entraîne une occupation accrue de l'ARN polymérase II à l'extrémité 5' de plusieurs unités de transcription, et exclusivement dans les cellules en réplication, ce qui suggère que la présence de la machinerie de réplication est nécessaire pour que cette accumulation se produise. Nous avons interprété cette

observation comme une conséquence de l'élimination manquée de l'ARN polymérase pendant les CTR. En effet, la densité de la polymérase le long des unités de transcription est inégale, avec un signal plus important au site de début de la transcription qui diminue progressivement vers la fin du gène, ce qu'implique que la machinerie de transcription semble être engagée plus longtemps à l'extrémité 5' par rapport à l'extrémité 3'. Dans un scénario purement stochastique, lorsqu'un réplisome entre dans une unité de transcription, il a donc plus de chances de rencontrer l'ARN polymérase à l'extrémité 5'. De plus, nous fournissons des preuves que la position des CTR change dynamiquement au cours de la réplication et, ainsi, l'accumulation observée de RNAPII à l'extrémité 5' des unités de transcription nécessite à la fois une réplication en cours et la co-présence physique au même endroit des machineries de transcription et de réplication.

En tournant notre attention sur de sites probables de CTR, nous montrons que la RFB ribosomale est un endroit utile et commode pour étudier la dynamique du processus. La rRFB est une forte barrière de fourche de réplication située dans la répétition ribosomique et qui permet la co-directionnalité entre la transcription et la réplication dans ce *locus* en bloquant la réplication de manière polaire grâce à la liaison de la protéine Fob1. L'absence de Sen1 au niveau du réplisome provoque une accumulation marquée de RNAPII à l'endroit où le réplisome est bloqué par la puissante barrière polaire mise en place par Fob1. Cette accumulation est située environ 100nt avant le site de liaison le plus fort de Fob1 (RFB1) et est donc compatible avec un scénario dans lequel le réplisome est bloqué et RNAPII s'accumule en amont et de façon co-directionnelle.

Nous avons également étudié le rôle des RNases H dans ces processus et l'impact des R-loops. A cette fin, nous avons développé une nouvelle méthode, H-CRAC, pour détecter les cibles RNaseH avec une résolution nucléotidique. L'étude des R-loops a reçu une grande attention ces dernières années de la part de la communauté scientifique. Cependant, les méthodologies actuellement utilisées pour évaluer les niveaux et la distribution des R-loops souffrent d'incohérences parfois marquées. Le pontage UV des RNases H à leurs cibles dans le H-CRAC s'est révélée être une méthode très efficace. Nous avons réussi à générer des cartes des cibles Rnh1, Rnh201 et aussi de l'enzyme humain, hsRNH1, exprimé chez la levure et montré qu'elles se chevauchent largement. Nous avons également mis en place plusieurs contrôles pour vérifier la fiabilité de notre approche, qui ont tous abouti à des résultats positifs confirmant que H-CRAC est une méthode valide pour cartographier les R-loops. Contrairement aux méthodologies basées sur le ChIP, H-CRAC offre une résolution sans précédent qui nous a permis de faire des observations intéressantes. En utilisant H-CRAC et

en réalisant des cartes de transcription dans les cellules délétées des RNases H, nous avons fait plusieurs observations inattendues. Une observation importante a été que la délétion de *RNH1* et *RNH201* phénocopie les effets de l'allèle *sen1-3*, en termes d'accumulation de ARN polymérase II à la rRFB. Ainsi, Sen1 et les RNases H sont tous deux nécessaires pour empêcher l'accumulation de la machinerie transcriptionnelle au niveau de la rRFB. Au contraire, le manque d'activité des RNases H n'a pas entraîné l'accumulation de RNAPII à l'extrémité 5' des autres unités de transcription.

Nous avons également découvert que Sen1 joue un rôle dans les conflits avec d'autres machineries. En fait, Sen1 interagit également avec l'ARN polymérase III, et très probablement en utilisant la même région de contact qu'avec le réplisome, puisque la liaison à l'ARN polymérase III est également perdue dans les cellules *sen1-3*. Cela nous a incités à surveiller l'occupation de l'ARN polymérase II à proximité des gènes d'ARNt. Nous avons observé une augmentation des signaux de l'ARN polymérase II à la fois en amont et en antisens des gènes d'ARNt dans les cellules *sen1-3*. Fait important, cet effet était indépendant de la réplication et n'était donc pas lié au rôle de Sen1 au niveau du réplisome. Nous observons également une occupation accrue de l'ARN polymérase II en antisens du gène ribosomal *RDN37*, où Sen1 et les RNases H semblent également enlever l'ARN polymérase II, peut-être pour éviter les conflits avec l'ARN polymérase I. Donc, nous proposons un modèle qui intègre les rôles des RNases H et Sen1 dans la résolution des CTR et sur d'autres sites de conflits. Ces résultats sont décrits dans un manuscrit en cours de soumission.

Les analyses préliminaires de nos cartes H-CRAC ont révélé une liaison importante de toutes les RNases H testées à l'extrémité 3' des gènes, près du site poly(A). La formation des R-loops dans une position similaire a déjà été observée dans des cellules de mammifère et suggéré chez la levure, bien que la résolution des données générées dans le passé soit trop faibles pour permettre des conclusions fiables. La présence d'un signal près du TTS est l'une des principales divergences entre S9.6-DRIP-seq et R-ChIP (les deux principales méthodologies actuellement utilisées pour mapper les R-loops genome-wide), le second étant généralement dépourvu d'un signal dans les régions de terminaison qu'il est au contraire trouvé par le premier. Ainsi, H-CRAC réconcilie pour la première fois ce qui a été observé avec les deux approches (c'est-à-dire S9.6 et la liaison sur l'ADN de RNase H1 or R-ChIP).

Nous avons effectué des analyses préliminaires de nos cartes H-CRAC afin de vérifier si les cibles des RNases H suivaient le comportement attendu d'après les résultats précédents et nous avons observé un certain accord avec la littérature. Par exemple, nous avons trouvé une

forte corrélation avec le taux de transcription, qui explique une grande partie de la variabilité observée dans le signal H-CRAC. Nous avons également observé un biais dans la composition nucléotidique des gènes formant un pic de RNase H à leur extrémité 3', en accord avec ce qui a été observé dans les cellules humaines.

Une caractéristique intéressante des gènes qui présentent un signal H-CRAC fort à l'extrémité 3' est la présence fréquente d'une unité de transcription convergente, ce qui pourrait induire des contraintes topologiques différentes par rapport aux gènes co-directionnels ou divergents. Dans le cadre d'une meilleure compréhension de la connexion entre Sen1, le signal en 3' et l'éventuelle contrainte topologique, nous avons découvert une interaction physique entre Sen1 et Top2 et obtenu le résultat inattendu que l'absence de Top1 et Top2 entraîne une mauvaise localisation de Sen1 et des défauts de terminaison de la transcription par l'ARN polymérase II. Ces derniers résultats sont préliminaires et discutés dans la partie finale du manuscrit.

TABLE OF CONTENTS

List of Figures and Tables

PROLOGUE

The DNA is home to a variety of molecular activities: transcription, replication, repair and chromatin remodelling are only some of a multitude of phenomena sharing the feature of involving DNA as a common ground for their motion.

In other words, the genome is far from being a quiet and calm field; on the contrary, a plethora of processes take place at the same time, and, surprisingly, in very close proximity, fostering an extremely crowded environment. Yet, all of this happens, and yet, evolution found a way to make it possible, ensuring that all those different tasks can be performed without interfering with one another. Such fine regulation relies on multiple mechanisms requiring the function of several actors. Pathological conditions in which such control fails to coordinate a proper environment surely helped us to appreciate its importance. A relevant example is provided by the coordination of transcription with replication, a major challenge in the organisation of the molecular processes taking place in the genome, and also the main subject of my work, here presented in this manuscript. Much evidence has been provided to support the notion that Transcription-Replication Conflicts (TRCs) are extremely detrimental for cell survival. Indeed, transcription can challenge the stability of replication forks, which when stalled become fragile elements underlying DNA damage.

Along the following chapters I will introduce the relevant literature useful for the well comprehension of the research carried out in the frame of my thesis work. In general, the description will focus on the *S. cerevisiae* model organism, but references to other species will be made when important for the elucidation of the process as well as for considerations about the conservation of some factors and mechanisms.

In the first part, I will discuss the most important facets of transcription, focusing on RNA polymerase II function. The second chapter will address how cells cope with the challenging task of duplicating their genome, in order to introduce the more relevant topics of topological

and replicative stress, fork stability and TRCs. Finally, a third chapter will be dedicated to R-loops, DNA:RNA hybrids that have been closely linked to TRCs, and to the enzymes involved in their metabolism.

INTRODUCTION

I. Transcription

In all cells, transcription is carried out by an RNA polymerase

The long path ensuring that the instructions stored in our genome can become useful for every cellular activity begins when an extremely large molecular machinery, the RNA Polymerase (RNAP), engages on the DNA template encoding for the desired information and catalyses the polymerisation of an RNA molecule, enchaining ribonucleotides complementary to the DNA stand.

Different kinds of RNAPs exist, but from bacteria to eukaryotes all of them share a conserved architecture of their core. In prokaryotes, all transcription is carried out by a single type of RNAP, whose core consists of four catalytic subunits (α, β, β' and ω), to which associates a single regulatory subunit known as sigma (s) to reconstitute the holoenzyme (Darst, 2001; Murakami and Darst, 2003).

Eukaryotes have three specialised RNAPs

Eukaryotic cells are, instead, equipped with three different kinds of RNAP, differing from one another for both their composition and the class of RNAs that they transcribe (Archambault and Friesen, 1993). RNA Polymerase I (RNAPI) is devoted to the expression of ribosomal RNAs; RNA Polymerase II (RNAPII) transcribes messenger RNAs (mRNAs) and some small regulatory RNAs; finally, RNA Polymerase III (RNAPIII) is mostly dedicated to the expression of tRNAs, the 5S rRNA and a few other RNAs. All three RNAPs are sophisticated enzymes, consisting of 14, 12, and 17 subunits, respectively, but they share several common features (Figure 1). The two largest subunits of all three eukaryotic RNAPs are unique but all related to the β and β′ subunits of the single bacterial RNAP. In addition, five subunits of the eukaryotic RNAPs are common to all three different enzymes.

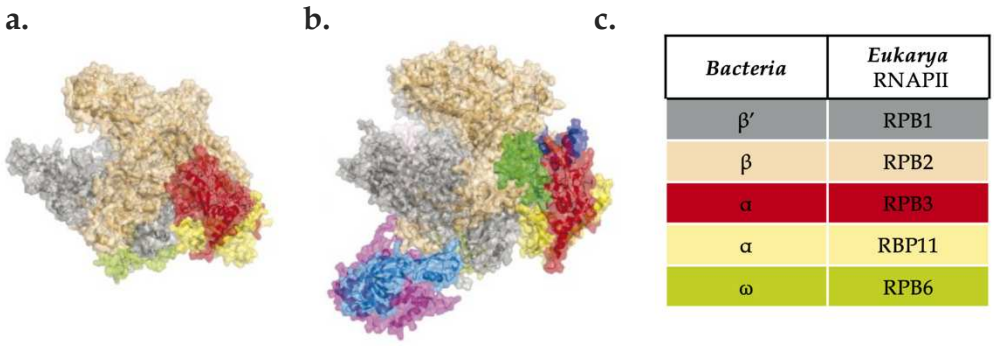


Figure 1. RNAPs structure across evolution.
(a) X-ray structure of the bacterial RNAP of *Thermus aquaticus* (PDB code: 1I6V). **(b)** X-ray structure of the eukaryotic RNA polymerase II of *Saccharomyces cerevisiae* (PDB code: 1NT9). **(c)** List of RNAP subunits with high conservancy across evolution. The colour code allows to identify the relative subunits in (a) and (b). *Adapted from Werner, 2008.*

RNAPII activity is highly regulated

Being deeply associated with gene expression and its regulation, RNAPII fulfils the most variegate tasks, and therefore it is not surprising that its structure and function have been thoroughly studied. The process of RNAPII transcription is tightly regulated, being gene expression essential to create complexity. *In vivo*, RNAPII requires a whole set of accessory factors orchestrating its function and establishing a transcription cycle that can be divided in three phases: initiation, elongation, and termination.

For matters of space and connections with the topic of this manuscript, the first two phases of the transcription cycle will be just briefly described, while more emphasis will be given to termination.

I.I Transcription initiation

The Pre-Initiation Complex assembles on a promoter

Transcription initiation of class II genes (RNAPII-dependent genes) starts with the assembly of the Pre-Initiation Complex (PIC), which is composed by a set of Transcription Factors (TFs) and by RNAPII in a non-active state. TFs can act either indirectly by recruiting enzymes that modify the chromatin structure, or directly by interacting with components of the transcription machinery. In the simplest model, both mechanisms collaborate to load RNAPII on a core promoter, the region of the DNA around the Transcription Start Site (TSS). Core promoters are not composed by a unique conserved sequence, instead they are a result of a diverse combination of often degenerated DNA modules, among which the TATA box, BRE (TFIIB-recognition element), Inr (Initiator element) and DPE (Downstream Promoter Element) are the most frequently found in metazoans (Hampsey, 1998) (Figure 2). Most promoters contain one or more of these elements, but no one among them is indispensable for promoter function.

Promoters are modular

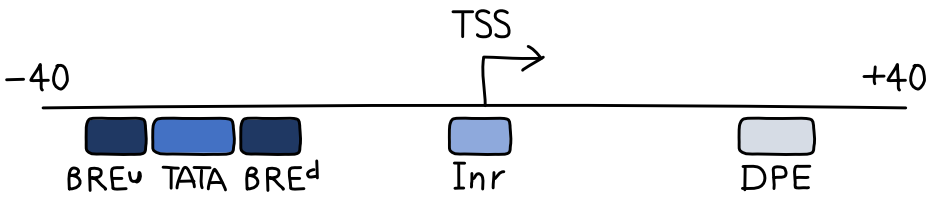


Figure 2. Most diffused core promoter elements in metazoans.
The position of the most diffused core promoter elements in metazoans relative to the Transcription Start Site (TSS) is schematised by the relative boxes roughly drawn to scale. The arrow indicates transcription directionality.

PIC assembly is a stepwise process involving GTFs

Formation of the PIC on a core promoter results from the stepwise assembly of General Transcription Factors (GTFs) with RNAPII (Figure 3). GTFs can be grouped into five main complexes: TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Struhl, 1995). The first step is the recruitment of TFIID (Ranish et al., 1999) via the sequence-specific binding of one of its components, the TATA-Binding Protein (TBP), to the TATA box element (Hernandez, 1993). TBP binds the widened minor groove, inducing a ~90°C bending of the DNA toward the major groove (Horikoshi et al., 1992). Interestingly, only about 10 to 30% of promoters, depending on the organism, contain canonical TATA boxes (Yang et al., 2007). Nevertheless, TBP is recruited to the vast majority of promoters (Kim and Iyer, 2004), most likely by nucleating protein-protein interactions among the GTFs and interacting non-specifically with DNA (Martinez et al., 1995). The second step consists in the addition of TFIIA and TFIIB, which both stabilise the binding of TFIID (Nikolov et al., 1995; Parvin and Sharp, 1993; Ranish et al., 1999). Then, loading of RNAPII with TFIIF leads to the formation of a more stable complex, named Core PIC (Chen et al., 2007; Eichner et al., 2010). The next stage is dedicated to the melting of the double helix, which is achieved through the DNA-dependent ATPase activity of TFIIH, stabilised by TFIIE, thus serving as a bridge with RNAPII (Goodrich and Tjian, 1994).

TATA-binding-protein starts PIC assembly

After DNA unwinding, the PIC moves to an open conformation, and transcription can start.

An NFR is essential to begin transcription

The DNA region comprising the core promoter is required to be unwrapped from nucleosomes to allow PIC assembly (Lai and Pugh, 2017). The creation of a Nucleosome Free Region (NFR) is achieved by multiple mechanisms, often relying on the recruitment of chromatin remodelers by gene-specific TFs. Moreover, efficient transcription activation relies on the function of the Mediator complex, which plays a role in the formation and stabilisation of the PIC as well as in acting as a bridge between gene-specific TFs and RNAPII.

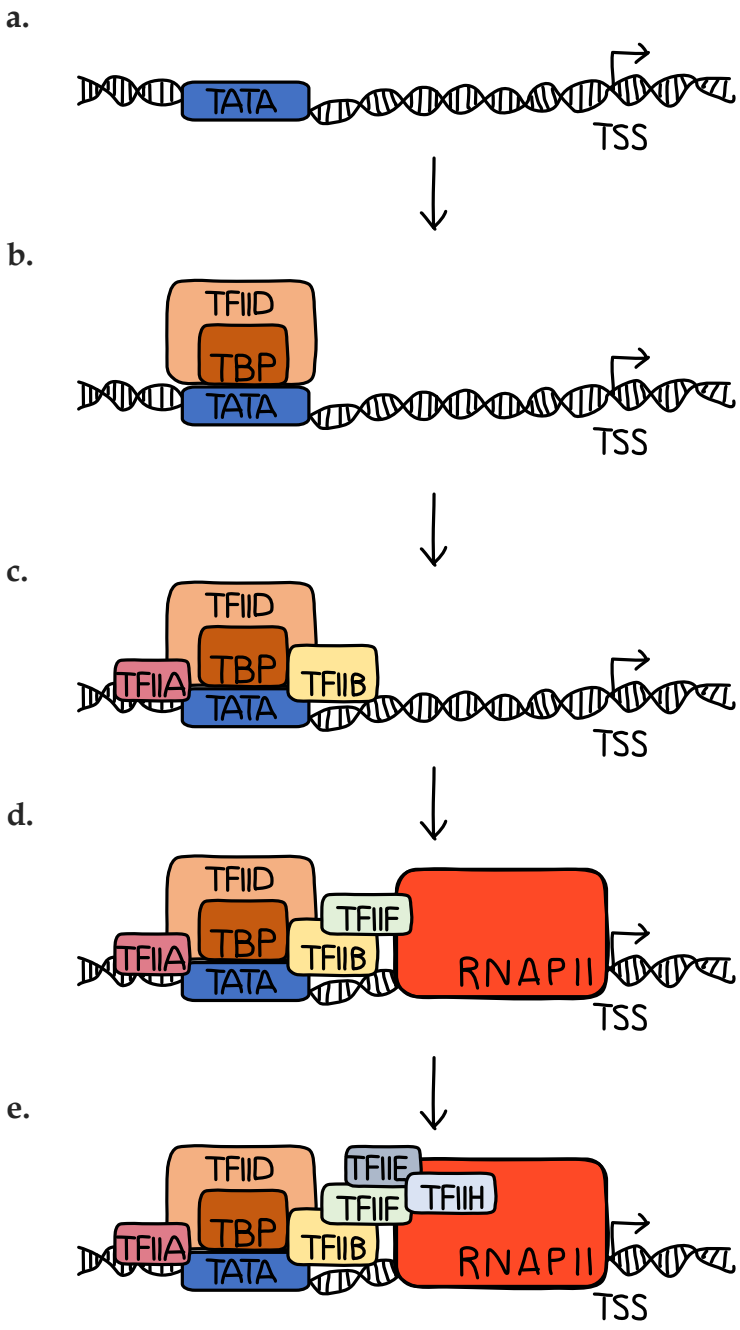


Figure 3. Stepwise assembly of the Pre-Initiation Complex.

(a) and **(b)** PIC assembly begins with TFIID recruitment via the sequence-specific binding of the TATA-Binding Protein (TBP). **(c)** The second step consists in the addition of TFIIA and TFIIB. **(d)** Then, loading of RNAPII with TFIIF leads to the formation of a more stable complex, named Core PIC. **(e)** The last step is dedicated to the melting of the double helix, which is achieved through the DNA-dependent ATPase activity of TFIIH, recruited via TFIIE.

I.II Transcription elongation

Promoter escape sets the onset of elongation and requires RNAPII dissociation from GTFs

When the PIC enters in an open state, RNAPII can proceed to the next phase of elongation. For this to happen, RNAPII must lose contacts with the GTFs associated with the core promoter, as well as start efficient polymerisation of the nascent RNA. From the synthesis of the first phosphodiester bond to the moment when the transcript reaches a length of about 15 nucleotides, the transcription machinery is unstable, and abortive transcription and promoter-proximal pausing are rather common at this stage (Holstege et al., 1997; Luse and Jacob, 1987), commonly referred to as promoter escape (Dvir, 2002). Several mechanisms contribute to the release of RNAPII from the promoter region. The dissociation of GTFs from the early transcription complex is particularly challenging and is promoted by both critical interactions within the polymerase and by the action of GTFs.

The CTD of RNAPII is extensively modified during transcription

Once the nascent RNA extends, the Elongation Complex (EC) is stabilised by the formation of a longer DNA-RNA hybrid within the holoenzyme, and after about 30 nucleotides the polymerase has adopted all the features of a proficient EC (Hahn, 2004). This phase coincides with extensive phosphorylation of the unstructured Carboxy-Terminal Domain (CTD) of the largest subunit (Rbp1) of RNAPII. Depending on the organism, the CTD is composed of a variable number of tandem repeats (26 in yeast, 52 in human) of the conserved heptapeptide Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7 (Harlen and Churchman, 2017). When recruited to the core promoter, the CTD of RNAPII lacks any modification, but already at the promoter escape stage phosphorylation begins to accompany the transcription cycle, with specific positions being phosphorylated and dephosphorylated when transitioning from one phase to the next, resulting in what can be defined as a CTD code (Buratowski, 2003; Egloff and Murphy, 2008; Komarnitsky et al., 2000) (Figure 4).

Several kinases mediate the phosphorylation of the RNAPII CTD. A relevant and well-studied example is the phosphorylation by the TFIIH Kin28 kinase (Cdk7 in human) on the serine 5 and 7 residues, which has

been shown to interfere with the interaction between Rpb1 and the Mediator, thus stimulating promoter escape (Jeronimo and Robert, 2014; Wong et al., 2014). Whereas Ser7-P is maintained up to the end of the Transcription Termination Site (TTS), Ser5-P is abruptly lost about 150 bp downstream of the TSS under the action of the Ssu72 and the Rtr1 phosphatases (Ganem et al., 2003; Hunter et al., 2016; Mosley et al., 2009) in favour of the Ser2-P, which increases later during transcription and is deposited by the Ctk1 kinase (Qiu et al., 2009).

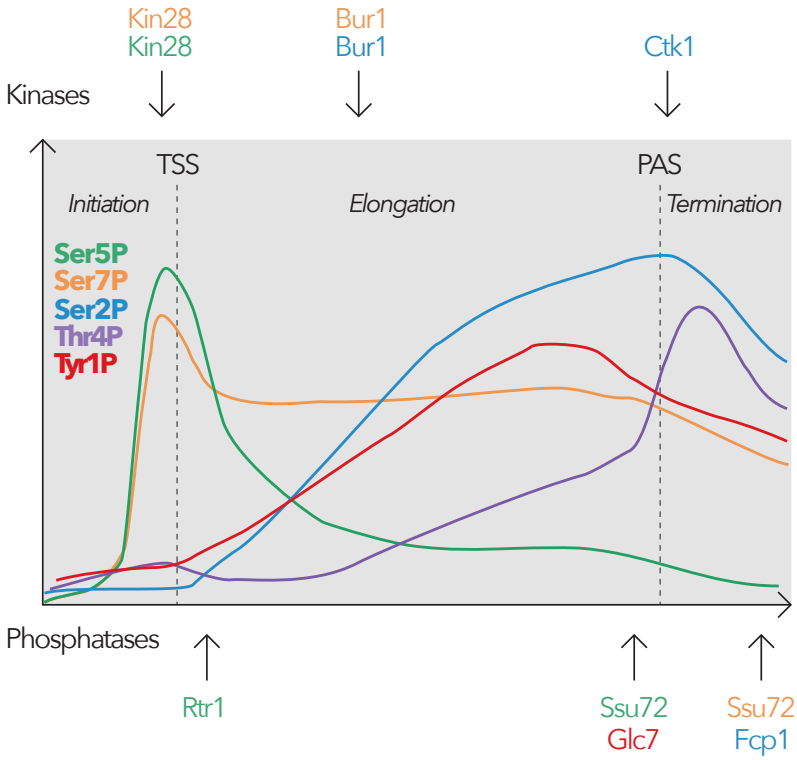


Figure 4. The CTD code along yeast protein-coding genes.
Schematic representation of the levels of Ser5, Ser7, Ser2, Thr4 and Tyr1 phosphorylation along a yeast coding gene. On the top and on the bottom of the scheme, kinases and phosphatases known to target these modifications have been indicated, with their position along the x axis roughly indicating their moment of action along the transcription cycle. The colour code follows the same used for each modification. *Adapted from Halen and Churchman, 2017.*

Phosphorylation of the CTD alters its affinity to co-transcriptional factors

The phosphorylation of each CTD residue activity alters its affinity for different components of the co-transcriptional machinery (Jeronimo et al., 2013), thus the dynamic pattern established by the CTD code couples transcription to RNA processing. For instance, Ser5-P is required for the binding of the 5'-capping enzymes, which ensures rapid protection of the nascent transcript (Komarnitsky et al., 2000; Schroeder et al., 2000). As described after, the same residue is also required for premature early termination by the Nrd1-Nab3-Sen1 complex (Kubicek et al., 2012; Vasiljeva et al., 2008). The Ser2-P, typical of the TTS region, is as well associated with the interaction to termination factors, as discussed in more details in the next section.

I.III Transcription termination

Termination requires the cleavage of the RNA and the dismantling of the EC

Termination is a vital step of the transcription cycle as it is required to both partition the genome, thus maintaining proper expression, and to recycle RNAPs allowing virtuous re-initiation. Efficient termination calls for at least three requirements: (i) the onset of transcription termination only at the end of the transcription unit, when a full-length RNA has been transcribed; (ii) the detachment of the nascent RNA from the polymerase to fulfil its fate; (iii) the dismantling of the elongation complex from the DNA template.

Two termination pathways coexist in yeast

In budding yeast, two different transcription termination pathways co-exist: the first one, mainly dedicated to the termination of mRNAs, relies on the function of the Cleavage and Polyadenylation Factor-Cleavage Factor (CPF-CF); the second one, which operates primarily on non-coding RNAs (ncRNAs), is instead carried on by the Nrd1-Nab3-Sen1 (NNS) complex. Both pathways are essential for cell survival, and share mechanistic similarities, yet have substantial differences, starting from the fate of the neo-synthetized transcripts.

I.III.I Cleavage and Polyadenylation Factor-Cleavage Factor termination

The CPF complex is a multisubunit termination factor

The main enzymatic actor of this pathway is the CPF-CF, a macromolecular complex composed of about 20 polypeptides in yeast, most of which evolutionary conserved (Kuehner et al., 2011). The CPF is composed by three modules: a polymerase, a phosphatase, and a nuclease module, to which two additional complexes, CFIA and CFIB, are associated and that are essentially involved in RNA recognition (Lidschreiber et al., 2018).

The poly(A) signal triggers termination

Termination by the CPF-CF comprehends two interconnected phases: the maturation of the 3'-end of the nascent transcript and the dismantling of the EC. The onset of termination is triggered by the presence of a specific sequence, a poly(A) signal, enclosed in the 3' Untranslated Region (UTR) of the nascent mRNA, and whose consensus is AAUAAA (Proudfoot, 2011) (Figure 5a). Despite being very well conserved in more complex eukaryotes, the requirement for a poly(A) signal is looser in budding yeast, with greater divergence from the consensus motif.

The RNA is cleaved, and a poly(A) tail is added to its 3'-end

Recruitment of CPF-CF to the EC depends on the interaction between the CFIA subunit Pcf11 and the Ser2-phosporylated CTD (Barillà et al., 2001; Kim et al., 2004a; Meinhart and Cramer, 2004). Once the poly(A) signal is embedded onto the nascent transcript it is recognised, together with additional sequence motifs, by different components of CFIA and IB (Rna15, Rna14, Hrp1, Cfp1, Cfp2) (Kyburz et al., 2003; Pancevac et al., 2010; Valentini et al., 1999), and the whole complex assembles on the RNA and cleaves it through the action of the endoribonuclease Ysh1 (CPF) (Kim et al., 2006; Mandel et al., 2006; Ryan et al., 2004) (Figure 5b). The release of the nascent RNA is coupled to the rapid addition of a stretch of adenosines to its newly formed 3'-end by the poly(A) polymerase Pap1 (PAP in human) (Amrani et al., 1997; Bienroth et al., 1993; Wahle, 1991). The addition of a poly(A) tail is required for protecting the RNA from degradation, to acquire competence for

nuclear export, which is indeed promoted also by its interaction with poly(A) binding proteins such as Nab2, and for efficient translation.

Two possible models coexist to explain how RNAPII is removed from the DNA

While proper positioning of the end of the transcript and release of the RNA are met by the presence of a specific signal at the end of the transcription unit and by the cleavage of the nascent transcript by Ysh1, the transcribing RNAPII still remains to be cleared from the DNA template. The mechanism by which this task is achieved is less clear and still to same extent matter of debate. Two not mutually exclusive models have been proposed (Porrua et al., 2016), both supported by independent findings.

The allosteric model

The first, known as allosteric model (Figure 5c), identifies in conformational changes the causal effect for RNAPII dissociation from the DNA template. More precisely, transcription of the poly(A) signal would cause conformational modifications leading to the combined loss of elongation factors in favour of the acquisition of termination factors to the polymerase, ultimately causing a destabilisation of the EC and its release from the DNA (Logan et al., 1987; Orozco et al., 2002). This model also implies that the cleavage of the nascent RNA is not a strict requirement for RNAPII dismantling. In support of this model, ChIP experiments have shown that the polymerase loses associated elongation factors before being released (Ahn et al., 2004; Kim et al., 2004a), while electron microscopy visualization of Miller's chromatin spreading in *X. laevis* and *D. melanogaster* unveiled that RNA cleavage occurs most frequently post-transcriptionally (Osheim et al., 1999, 2002). Moreover, *in vitro* experiments with purified elements from both *S. cerevisiae* and *D. melanogaster* revealed that Pcf11 can destabilize the EC by simultaneous binding to the RNAPII CTD and to the nascent RNA (Zhang and Gilmour, 2006; Zhang et al., 2005). On the same line, it was more recently shown that cleavage of the RNA is not a requirement for RNAPII termination in mammalian *in vitro* reconstituted systems (Zhang et al., 2015).

The torpedo model

The second model, named torpedo model (Figure 5c), implicates the action of an exonuclease that takes advantage of the formation an uncapped 5'-end after the co-transcriptional cleavage of the nascent

RNA to employ it as an entry site and progressively digest the RNA to finally reach and displace RNAPII from the DNA. Pivotal studies supporting this model showed that the yeast 5′-3′ exonuclease Rat1 and its human homolog, Xrn2, play a role in RNAPII termination (Kim et al., 2004b; West et al., 2004). Indeed, in both cases, it was demonstrated by ChIP experiments that inactivation of the exonucleases results in transcriptional read-through of RNAPII after the termination site, which was also confirmed at the genome-wide level more recently (Baejen et al., 2017; Eaton et al., 2018; Fong et al., 2015).

In yeast the Rat1 nuclease co-purifies in complex with Rai1 and Rtt103 (Kim et al., 2004b). The former functions as a co-factor for Rat1, and its deletion causes termination defects. The latter is a scaffold protein that bridges Rat1-Rai1 to the EC via its interaction with the Ser2-P residue of the CTD. Nevertheless, it has been shown that neither Rat1 (alone or in complex with its cofactor Rai1) nor degradation of the nascent RNA by itself is sufficient to elicit termination in a highly purified *in vitro* assay (Dengl and Cramer, 2009), while it was the case in less purified ECs from whole cell extracts (Pearson and Moore, 2013), suggesting that additional factors are necessary to achieve termination. Importantly, addition of Rtt103 to an *in vitro* termination reaction allows RNAPII removal when using an exonucleolytically inactive mutant of Rat1 (Lunde et al., 2010; Luo et al., 2006; West et al., 2004). This led to the proposal that Rat1 (in complex or not with Rai1) needs to catch up with the polymerase to exert its function in dislodging the EC, maybe by inducing allosteric changes in the polymerase. The RNA would represent a 'route' for Rat1 to reach RNAPII, but simple tethering of the protein to the polymerase by Rtt103 would then compensate for the absence of the RNA (i.e., in a catalytically inactive Rat1).

Pausing of the polymerase has been shown to be important, albeit not sufficient, for effective termination, especially in metazoans (Core et al., 2008; Gromak et al., 2006; Nag et al., 2007). The presence of RNAPII pausing at 3′-end positively correlates with the strength of the poly(A) site, but it does not depend on the CTD. In this scenario, slowing of

RNAPII allows Rat1 or Xrn2 to quickly catch up with the polymerase to exert its function.

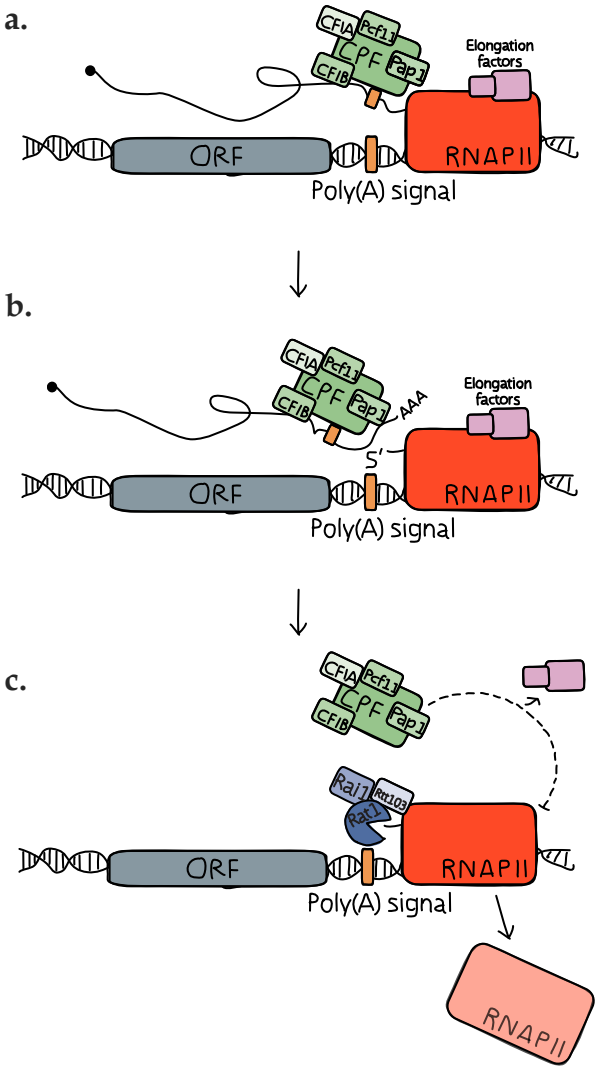


Figure 5. The CPF-CF termination pathway in budding yeast.
(a) The CPF-CF complex is recruited to the nascent transcript by interaction with the poly(A) signal. **(b)** The complex cleaves the RNA and promotes the addition of a poly(A) tail to the nascent RNA via the action of the poly(A) polymerase Pap1. **(c)** Removal of RNAPII from the DNA template via the torpedo (Rat1-Rai1-Rtt103) and the allosteric (CPF-CF) model.

Finally, because of these contradictory and sometime hard to reconcile findings, a combined model has been proposed, positing that *in vivo* the

allosteric and torpedo model act in concert to promote transcription termination (Luo et al., 2006).

I.III.II The NNS pathway

As mentioned before, the second termination pathway in yeast is mostly dedicated to the termination of ncRNAs. Aside from the actors involved, this pathway differs from the poly(A) termination for the fact that it is coupled to the degradation machinery that completely or partially digests, according to their nature, the produced transcripts.

The NNS complex is composed by the two sequence-specific RNA binding proteins Nrd1 and Nab3 and by the evolutionary conserved helicase Sen1 (SETX in human). All three subunits of the complex are essential in budding yeast, but Sen1 is the sole with a catalytic activity.

Nrd1 is an essential RNA-binding protein and it interacts with the CTD of RNAPII

Nrd1 is a 64 kDa protein with homology to the mammalian SCAF8 and SCAF4 anti-terminator factors (Yuryev et al., 1996). All these proteins hold an N-terminal CTD-Interacting Domain (CID), a single RNA recognition motif (RRM) and a segment enriched in alternating charged residues, containing several arginine-glutamate and arginine-serine (RE/RS) dipeptides (Steinmetz and Brow, 1998). The CID domain of Nrd1 preferentially interacts with the Ser5-phosphorylated CTD of RNAPII, which is typically prominent at the beginning of the transcription cycle (Kubicek et al., 2012; Vasiljeva et al., 2008). This interaction favours an early recruitment of the NNS complex to transcribing RNAPII, and it is important for efficient termination, even though not strictly required (Han et al., 2020; Tudek et al., 2014).

Nab3 forms a heterodimer with Nrd1

Nab3 is a 90 kDa acidic ribonucleoprotein containing a RRM motif, flanked by two low-complexity domains. The N-terminal domain is an aspartic/glutamic acid rich region (D/E rich) dispensable for viability and its function is unknown (Loya et al., 2012, 2013). The C-terminal domain is a glutamine/proline rich region (Q/P rich) and, on the

contrary to the N-terminal domain, is essential and might play important roles in Nab3 self-association. Over-expression of Nab3 suppresses some *nrd1* temperature-sensitive (*ts*) alleles (Conrad et al., 2000), suggesting that they function similarly. Indeed, Nrd1 and Nab3 have been shown by both genetic and biochemical approaches to form a tight heterodimer (Carroll et al., 2007; Conrad et al., 2000), which allows a cooperative binding of the two proteins to their consensus sequences (GUAA/G and UCUUG, for Nrd1 and Nab3 respectively) (Carroll et al., 2004; Conrad et al., 2000; Hobor et al., 2011; Porrua et al., 2012; Steinmetz and Brow, 1998).

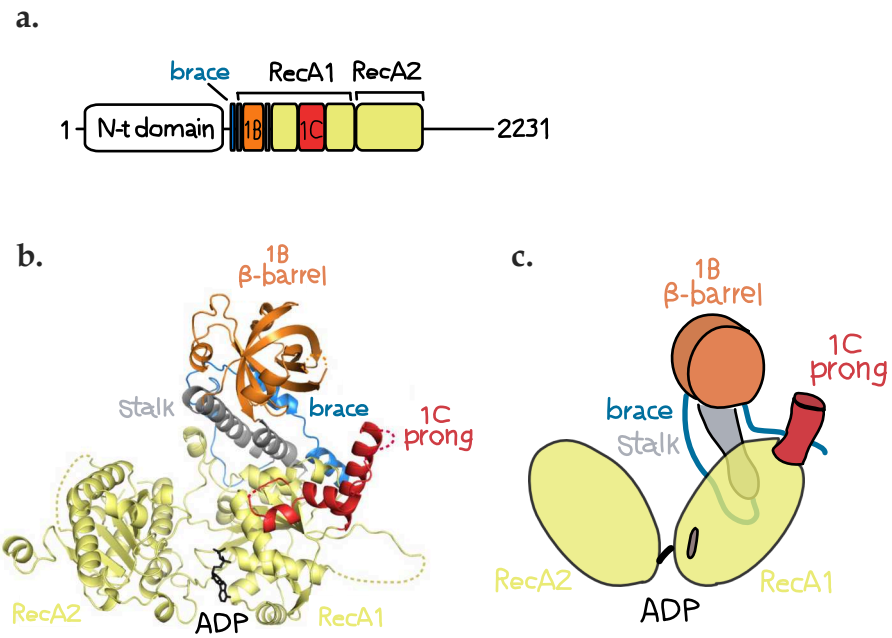


Figure 6. Structure of the helicase domain of Sen1.

(a) Schematic representation of the structured modules (boxes) and of the unstructured regions (lines) of Sen1. **(b)** Structure of the helicase domain of Sen1 in absence of RNA. **(c)** Simplified representation of the major structural elements of Sen1. The same colour code is used in the whole figure. *Adapted from Leonaité et al., 2017.*

Sen1 is an essential and conserved Upf1-like helicase

Sen1 is a 252 kDa RNA and DNA helicase belonging to the Super Family IB (SFIB) Upf1-like group of the helicases (Figure 6). It consists of a large N-terminal domain proposed to mediate interaction with RNAPII, Rnt1 and Rad52 (Finkel et al., 2010; Ursic et al., 2004); a central conserved helicase domain that contains two RecA modules, with the classical helicase motifs involved in nucleic acid binding and ATP hydrolysis; finally, a C-terminal region without any predicted secondary structure that interacts with Nab3 and Glc7 (Nedea et al., 2008).

Sen1 is the sole catalytic component of the NNS complex

Several studies have allowed to characterise a quite detailed mechanism by which the NNS complex functions (Figure 7). First, *in vitro* experiments have shown that Sen1 is responsible for the dismantling of the EC complex in an ATP-dependent manner (Porrua and Libri, 2013). Importantly, a Sen1 truncated protein composed only of its helicase domain is sufficient to recapitulate the activity of the full-length enzyme *in vitro* (Leonaitė et al., 2017). However, deletion of the N-terminal domain of Sen1 causes lethality (Han et al., 2020), suggesting that the mere catalytic activity is not sufficient for effective termination *in vivo*. The contribution of Nrd1 and Nab3 is still debated, but the two proteins might contribute to recruit Sen1 to its targets by interacting with their binding sites. Indeed, Sen1 must translocate 5'-3' on the nascent RNA to enter in contact with RNAPII, in order to dissociate the EC from the DNA by both exerting a mechanical action and inducing conformational changes that were recently observed by single molecule studies (Wang et al., 2019).

Late crystallographic studies have shed light on Sen1 structure showing that from the first RecA module emerge the 'stalk', the 'prong' and the 'barrel' auxiliary domains (Figure 6) (Leonaitė et al., 2017). In general, the core domain is conserved among related helicases, however the accessory domains define specific features unique to Sen1 that underlie its ability to accomplish RNAPII removal. Notably, the prong domain, which is essential for termination, has a flexible nature that could facilitate its insertion in the RNAPII exit channel, resulting in profound conformational changes and destabilization of the EC.

The nuclear exosome degrades or processes RNAs terminated by the NNS complex

As stated above, a unique feature of the NNS pathway is to be coupled to decay or processing of the RNA. The nuclear exosome is a large conserved multisubunit complex endowed with 3'-5' exonuclease and endonuclease activities accounting for the vast majority of RNA nuclear degradation as well as for the processing and maturation of functional non-coding RNAs (Chlebowski et al., 2013; Mitchell et al., 1997). It exists also in a cytoplasmatic version, which contains a common core and associated specific subunits. The core exosome is composed of nine rather small polypeptides structurally organised in a toroidal architecture reminiscent of some prokaryotic complexes involved in RNA metabolism. The core exosome is associated with other two subunits, Dis3 and Rrp6, with the second one being exclusively nuclear.

The TRAMP complex links NNS termination to exosome processing

The Trf–Air–Mtr4 polyadenylation (or TRAMP) complex is alternatively composed of a poly(A) polymerase (Trf4 or Trf5), a zinc knuckle RNA binding protein (Air1 or Air2) and the DExH-box RNA helicase Mtr4 (LaCava et al., 2005; Vaňáčová et al., 2005; Wyers et al., 2005). TRAMP is an important cofactor of the exosome, as it aids degradation by unstructuring the RNA through the Mtr4 helicase and by adding a short poly(A) tail to the 3'-end of the RNA, which strongly enhances the affinity of the exosome for the molecule (Jia et al., 2012).

After the dismantling of RNAPII from the DNA, Nrd1 and Nab3 remain post-transcriptionally associated to the RNA. Nrd1 recruits TRAMP through the direct recognition of a CTD mimic, known as Nrd1-interacting motif (NIM), in the TRAMP component Trf4 (Tudek et al., 2014). The sequential (and mutually exclusive) interaction of Nrd1 with the CTD and Trf4 contributes to the temporal coordination of termination with degradation. After addition of a poly(A) tail by the TRAMP complex to the terminated transcript, its rapid degradation or processing is ensured by multiple interactions between the exosome and both TRAMP and NNS components.

As already mentioned, the exosome can either completely or partially digest the terminated RNA. The choice between these two different fates is tightly linked with the nature of the RNA molecule. Products of pervasive transcription, such us Cryptic Unstable Transcripts (CUTs)

(described later, see Introduction, § I.IV), are completely digested, and this was recently shown to facilitate the recycling of the Nrd1-Nab3 heterodimer, otherwise sequestered by the RNA (Villa et al., 2020). Functional non-coding RNAs, such as snoRNAs, are instead trimmed by the exosome up to the size of the mature product, likely because the snoRNP core ribonucleoprotein complex prevents further degradation of the molecule.

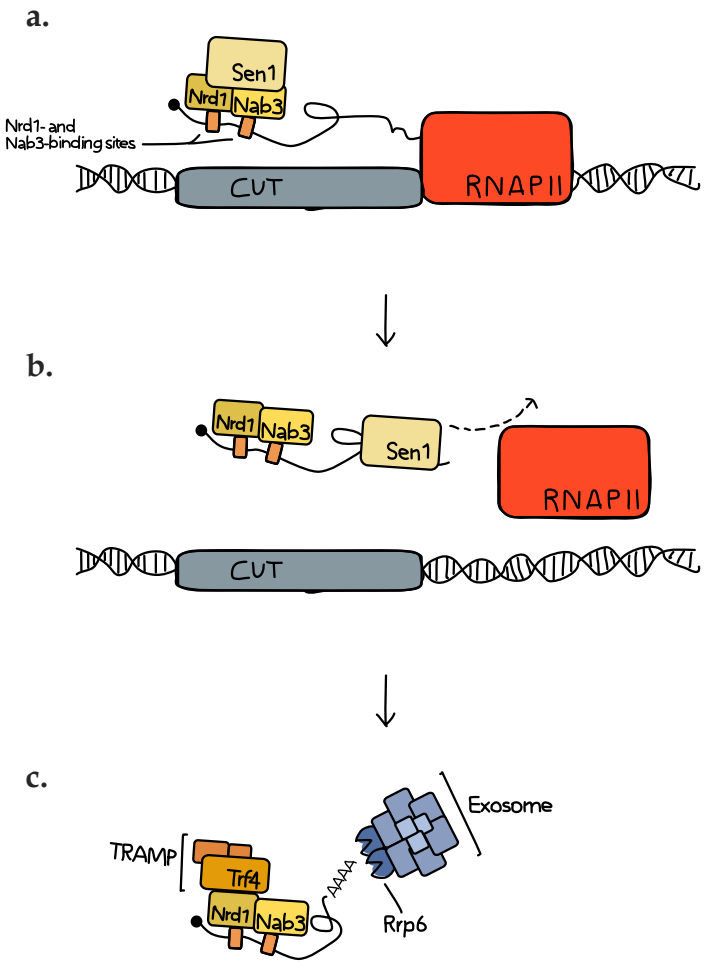


Figure 7. Termination of RNAPII at non-coding RNAs via the NNS complex.
(a) Recruitment of the elongation complex via recognition by Nrd1 and Nab3 of specific motifs embedded into the nascent non-coding RNA. **(b)** Removal of the elongation complex through the ATP-dependent helicase activity of Sen1. **(c)** The non-coding RNA is targeted to processing or complete degradation via the combined action of the TRAMP complex and of the exosome.

I.IV Pervasive transcription

Genomes are pervasively transcribed by RNAPII

On the contrary of what has been thought for several years, the transcription processes are not restricted to regions encoding information to produce mRNAs or other functional RNAs. Indeed, genomes are instead pervasively transcribed, with active RNAPII found virtually everywhere. About 15 years ago, a collaborative work from the Libri, Jacquier and Seraphin laboratories reported the existence of a large class of non-coding RNAs originating from intergenic or antisense regions of the genome, which, because of their instable nature, were named Cryptic Unstable Transcripts (CUTs) (Wyers et al., 2005). These transcripts are so quickly degraded after being transcribed that they cannot be detected in wild-type cells and their discovery was possible only in conditions in which their decay was impeded (Davis and Ares, 2006; Houalla et al., 2006; Wyers et al., 2005).

CUTs are the most abundant non-coding transcripts, and they are terminated by the NNS

During the past decade, other classes on non-coding transcripts have been found, and they differ from each other mostly for their termination pathway, and consequently for their stability. CUTs are the most abundant class. They are transcribed by RNAPII, and therefore endowed with a cap at their 5′-end. Their length is usually between 200 and 500 nt, and they are typically targeted by the NNS pathway (Arigo et al., 2006; Thiebaut et al., 2006), which ensures their rapid clearance via the action of the nuclear exosome. Accordingly, deletion of the non-essential *RRP6* gene, encoding for the nuclear exosome nuclease, leads to a massive accumulation of these molecules in the nucleus (Gudipati et al., 2012; Wyers et al., 2005).

SUTs and XUTs are terminated by CPF-CF and degraded by Xrn1

Stable Unannotated Transcripts (SUTs) are also transcribed by RNAPII, but they are longer and more stable compared to CUTs (Xu et al., 2009). Indeed, they are terminated by the CPF-CF complex, even though the NNS complex might also partially contribute, and therefore their degradation depends to a lesser extent on the exosome (Gudipati et al., 2012; Marquardt et al., 2011). Instead, they are exported to the cytoplasm where they are primarily degraded by the Xrn1 5′-3′

exonuclease (Marquardt et al., 2011). Xrn1-sensitive Unstable Transcripts (XUTs) are a third class of RNAPII non-coding transcripts, and they are similar to SUTs as they are also exported to the cytoplasm and, as their name suggests, degraded by Xrn1 (van Dijk et al., 2011). As a matter of fact, they differ mostly for their stability, being rather unstable and undetectable in a wild-type context. It is important to note that, however, the distinction among these different categories is sometimes blurry: a certain transcript can be surely predominantly targeted by one of the termination pathways (which, as already mentioned, is the main determinant for the category to which a non-coding RNA belongs), but as biological processes are not fully efficient, its termination can also depend on other pathways. For example, a fraction of the transcription events of a CUT could escape the NNS pathway and rather be terminated to a downstream poly(A) site by the CPF-CF complex. Those RNAs would then be exported from the nucleus and depending on their stability classified as XUTs or SUTs.

Table 1. Transcript categories in yeast and their termination and processing pathways.

Transcript	Termination pathway	Stability	Degradation factors
mRNA	CPF–CF	Stable	None
snRNA and snoRNA	NNS; Pcf11	Stable (3′-end processed)	TRAMP, Rrp6, exosome, Rex1 (3′-end processing)
CUT	NNS	Unstable	TRAMP, Rrp6, exosome
SUT	CPF–CF and possibly NNS	Partially unstable	Rrp6, exosome, Xrn1 (NMD)
XUT	CPF–CF	Unstable	Xrn1 (NMD)
RUT	Reb1 roadblock	Unstable	TRAMP, Rrp6, exosome

ncRNAs originate from NFRs

The requirement for an NFR to initiate transcription is a general rule that also applies to cryptic RNAs, which in fact originate mostly from 3′ or 5′ regions of genes, naturally depleted of nucleosomes, or from

cryptic promoters in intergenic regions (Malabat et al., 2015; Neil et al., 2009). Strikingly, many cryptic RNAs stem from divergent transcription of genes, a feature that fits with the notion that promoters are intrinsically bidirectional.

Multiple mechanisms control pervasive transcription

Many cryptic transcripts are produced in the proximity of coding genes, that can be overlapped in the antisense or the sense direction. As a consequence, it is extremely important to control pervasive transcription (Jensen et al., 2013). Many observations have indeed proved that failure to control pervasive transcription poses a real danger for cell survival. First, lack of degradation of cryptic transcripts leads to their accumulation in the nucleus, where they are susceptible to titrate processing and export factors. This is prevented by the already mentioned degradation pathways. Second, if pervasive transcription is not properly terminated or its rate of transcription is not kept low, it can interfere with the transcription of neighbouring coding genes, through a mechanism known as 'transcriptional interference' (Shearwin et al., 2005). Non-coding transcription through the promoter of a nearby gene can disturb the expression of the coding gene by challenging the binding of TFs to the downstream promoter or by altering the epigenetic status of the nucleosomes in this region. Nucleosomes deposited after an elongating RNAPII are enriched of H3K4me2 and H3K36me, modifications that mark the body of the gene and are repressive for transcription initiation (Kim et al., 2016, 2012; Nevers et al., 2018). Therefore, an unterminated polymerase running from an upstream ncRNA might redefine the downstream promoter of a coding gene as a 'gene body' region, preventing transcription initiation from the canonical TSS. The chimeric transcripts will be very unlikely to bypass the RNA quality controls and consequently degraded. Transcriptional interference has been shown to occur genome-wide upon depletion of NNS subunits, leading to major changes in the transcriptome and thus explaining their essentiality (Schaughency et al., 2014; Schulz et al., 2013).

Pervasive transcription can interfere with transcription of mRNAs and thus regulate gene expression

Transcriptional interference is so far the only mechanism by which non-coding transcription has been clearly shown to exert a function in yeast,

such as for the expression of some genes belonging to the pathways of nucleotides synthesis (Kuehner and Brow, 2008; Thiebaut et al., 2008). On the contrary, ncRNAs appear to lack well-defined regulatory-functions and have poor coding potential, which strongly questions their role.

II. Replication

For life to propagate cells must divide, and surely one of the most challenging tasks to bypass to reach this goal is to generate a duplicate of all the information contained in the genome.

Many mechanisms at the basis of the process of replication are conserved from bacteria to humans

DNA replication is an exceptionally intricate process that exploits the complementarity between DNA strands to synthetize a brand-new copy of the genome and ultimately guarantee that each daughter cell will inherit its own DNA after division. In such a complex phenomenon several factors are implicated, and many demanding efforts must be undertaken. General rules and common mechanisms can be found from the simplest to the most complex organism. For instance, the process of copying must be as flawless as possible, so that the next generation will receive a nearly identical genome; in every organism replication takes place in a specific temporal moment, known as S-phase, during which DNA synthesis is finalised before progressing to the next stage of the cell cycle; the polymerisation of the newly synthetized strands is performed by a multisubunit machinery named replisome, the overall composition of which is extremely similar and well conserved although specific components may differ from one organism to the other.

Akin to transcription, also replication can be divided into three stages: initiation, elongation, and termination.

II.II Replication initiation

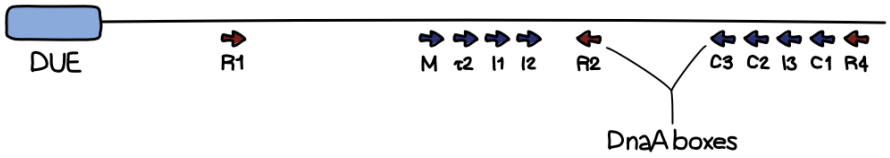
"Once and only once"

Each sequence contained in the genome must be replicated once and only once. If not, the next generation will be provided with either more copies of the same sequence, causing dosage misbalancing, or either will lack some of the parental genetic information. In both cases the consequences can be catastrophic, as reminded us by many human pathologies in which this rule fails to be complied.

During licensing, the pre-RC complex assembles on an ORI

The processes involved in the respect of this basic principle starts before the synthesis of the new copy of the DNA, during the G_1 phase of the cell cycle. At this stage, specific sequences along the chromosome, named origins of replications (ORIs) serve as a platform for the assembly of a pre-Replication Complex (pre-RC) (Bell and Kaguni, 2013; Jacob et al., 1963; Remus and Diffley, 2009).

a.



b.

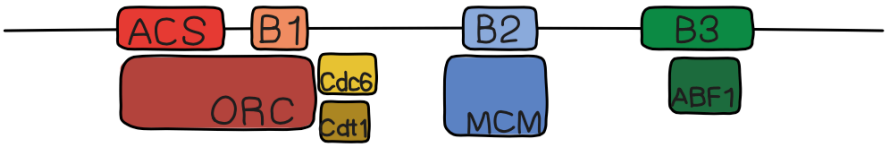


Figure 8. Composition of bacterial and budding yeast ORIs.
(a) Schematic representation of the *E. coli* origin (*oriC*). The arrows indicate the DnaA recognition boxes, and their colour designates their affinity (red, high; blue, low). **(b)** Schematic representation of the *S. cerevisiae* origin of replication and of its elements. The main proteins that bind to each element are also presented.

In bacteria a single *oriC* ensures replication of the chromosome

In almost all the prokaryotes, where a single circular chromosome is usually found, an individual ORI, named *oriC* (Figure 8a), fires two replication forks in the opposite orientation which will then meet in a terminator region to complete the duplication of the whole genome. *OriCs* are most commonly found in the immediate proximity of the region encoding for DnaA, the bacterial initiator protein (Gao and Zhang, 2007; Mackiewicz et al., 2004), and as a general rule, they are generally constituted of an extremely A-T-rich region named DNA

DnaA is the bacterial initiator protein

Unwinding Element (DUE) (Kowalski and Eddy, 1989) that facilitates DNA melting and by a variable number of 9-bp motifs that are recognized by DnaA, known as DnaA boxes, with different binding affinities (Figure 8a). DnaA belongs to the AAA+ family of proteins, ATP-associated proteins with different functions in the cell (Katayama et al., 2010; Nievera et al., 2006; Schaper and Messer, 1995). The ATP binding and hydrolysis creates an alternative state of the protein which determines its activity and is important for ensuring that the pre-RC fires only once per cell cycle. As discussed later, the same basic mechanism is also found in eukaryotes. *OriC* from *E. coli* and other bacteria also contains binding sites for modulators that can influence DnaA interaction with its boxes, and that can be both positive (activators) and negative (repressors).

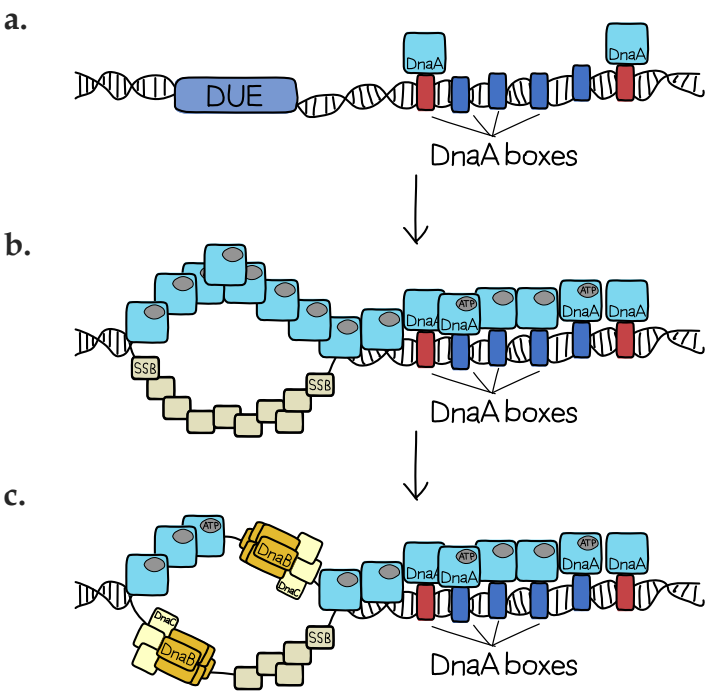


Figure 9. Origin licensing in bacteria.

(a) The high affinity sites (red boxes) are always bound by DnaA. **(b)** At replication onset DnaA-ATP binds also the low affinity boxes and it oligomerises along the origin melting the

DNA at the DUE. **(c)** The replicative helicase DnaB and its loader DnaC are assembled at the origin.

ATP regulates DnaA activity

The temporal control of replication initiation is based on the different affinity of DnaA for its binding sites (Figure 9). The high affinity binding sites are almost always bound by DnaA regardless of its ATP state, while the lower affinity sites are recognised only at replication onset and exclusively by the ATP-bound DnaA (McGarry et al., 2004). As DnaA oligomerises along the origin, it promotes the formation of a large nucleoprotein complex that facilitates the DNA melting within the DUE, allowing the positioning of the replicative helicase DnaB and its loader DnaC (Marszalek and Kaguni, 1994; Sutton et al., 1998). Regulatory mechanisms subsequently inactivate DnaA by stimulating ATP hydrolysis upon replisome assembly, thus ensuring that the pre-RC cannot be reassembled until the next cell cycle (Katayama et al., 1998).

Multiple origins are found along eukaryotic chromosomes but not all of them are activated

Eukaryotes have much more complex genomes than bacteria, with a variable number of linear long chromosomes and a one only ORI would not be sufficient for ensuring duplication of the whole genome. A single human chromosome would take about a month to be fully replicated from an individual ORI; a timing hardly compatible with life. Therefore, multiple ORIs are scattered along eukaryotic chromosomes so that replication can simultaneously start from many spots and ensure rapid completion of the process. In budding yeast, about 400 ORIs are distributed along the sixteen chromosomes (Siow et al., 2012), whereas in human about 50.000 to 80.000 origins are estimated to be present (Besnard et al., 2012; Cadoret et al., 2008; Mesner et al., 2011). In every case, only a subset of these origins is actually activated in each cell cycle, and the group of activated origins is variable from one cell cycle to the next or from one cell type to the other. Only one out of five origins is used per each replicon and the choice among them is apparently stochastic, yet influenced by other factors such as nearby transcription and the chromatin state.

Budding yeast origins are determined by a sequence motif known as ACS

S. cerevisiae ORIs are the best characterised ones, also because this is so far the only eukaryote for which it was possible to clearly demonstrate that a consensus motif specifies an origin. These sequences were identified thanks to their ability to confer autonomous replication to episomes, hence they were designated Autonomously Replicating Sequences (ARSs) (Stinchcomb et al., 1979). ARS function depends on the co-presence of 3 distinguished modules: namely A, B and C domain (Figure 8b). The A domain is composed of the 11 bp sequence (A/T)TTTA(T/C)(A/G)TTT(A/T), known as Autonomous Consensus Sequence (ACS) (Broach et al., 1983; Dhar et al., 2012; Theis and Newlon, 1997), and recognised by the initiator protein, the evolutionary conserved heterohexameric Origin Recognition Complex (ORC) (Bell and Stillman, 1992). The B domain is a 100 bp *cis*-acting element located downstream from the ACS that contains DUEs and binding sites for ORC components, the replicative helicases and Abf1 (ARS binding factor 1), that plays a role in transcription and can influence chromatin structure and nucleosome assembly (Buchman et al., 1988; Huang and Kowalski, 1993; Lipford and Bell, 2001; Rao and Stillman, 1995; Rowley et al., 1995). More recently, B elements have been proposed to actually be reverse A domains with lower affinity where a second ORC binds in an opposite orientation and thus establishing bidirectionality (Coster and Diffley, 2017). The C domain is placed upstream the ACS and holds binding sites for certain transcription factors that stimulate, but are not essential, for the origin activity (Chang et al., 2008; Lascaris et al., 1999; Lynch et al., 2005). Another important requirement for ARS function is that it must be depleted of nucleosomes. This is achieved by the intrinsic propensity of A-T polymers to exclude nucleosomes, but also by the ORC binding itself, which induces a periodic positioning of the nucleosomes adjacent to the ACS (Eaton et al., 2010).

Metazoan ORIs are not determined by a sequence

In more complex eukaryotes, including humans, the specific features defining an ORI have been more difficult to characterise for different reasons. First, experiments in which random sequences are inserted in plasmids in order to identify sequences conferring competence to replicate, as was done in yeast, is challenged by the general poor efficiency of replication in these systems. Second, no clear consensus

sequence appears to mediate the specification of the origins, challenging their identifications until the development of genome-wide techniques (Besnard et al., 2012; Cadoret et al., 2008; Mesner et al., 2011).

G4 structures, chromatin, and transcription influence ORIs in metazoans

Despite the lack of a clear consensus motif, the analysis of tens of thousands of replication origins showed a certain tendency for the presence of G-rich sequences with the potential of forming a G quadruplex (G4), a noncanonical four-stranded helical structure (Rhodes and Lipps, 2015), which has been proposed to help creating a favourable nucleosome positioning for replication initiation and also to be recognised by initiator proteins (Cayrou et al., 2011, 2015). However, the frequency of sequences prone to form G4 in the human genome is largely above the number of ORIs, entailing that a G4 is not sufficient to promote replication initiation (Huppert and Balasubramanian, 2007). To complicate the picture, more recent studies have revealed a contribution for both epigenetics modifications and transcription in the determination of ORIs (Sequeira-Mendes and Gómez, 2012). For instance, many origins have been shown to coincide with active promoter elements at CpG islands (Cadoret et al., 2008; Delgado et al., 1998; Sequeira-Mendes et al., 2009).

The ORC complex is the eukaryotic initiation protein

The ORC complex is composed by six subunits (Orc1-6), five of which are predicted to belong to the AAA^+ family of proteins (Li and Stillman, 2012). Similarly to DnaA binding to its high affinity sites, ORC also interacts with the ORIs almost always throughout the cell cycle (Diffley and Cocker, 1992; Diffley et al., 1994). The ORC structures as a ring with a gap from which the DNA is introduced into the ORC central channel (Li et al., 2015b). After DNA binding, the gap is sealed by the co-loader Cdc6 in an ATP-dependent manner, trapping the DNA in the centre of the ORC-Cdc6 toroid (Speck et al., 2005) (Figure 10a). The subsequent step is the loading of the MCM helicases in a non-active state, which is accomplished by the concerted action of Cdc6 and Cdt1 (Figure 10b). The Mini Chromosome Maintenance (MCM) helicase is a conserved heteromeric complex found in all sequenced eukaryotes and constituted by six distinct but evolutionary related Mcm (Mcm2-7) proteins, which all contain an ATPase domain at their C-terminal end (Chong et al., 1996). Two non-active MCM hexamers are loaded as a

Two MCM helicases are loaded in a bidirectional manner at each ORI

ring around dsDNA at each ORI in an ATP-dependent manner and in a two steps-mechanism, but it not clear yet if this requires one or two ORC complexes (Coster et al., 2014; Kang et al., 2014). The direct consequence of ATP hydrolysis during helicase loading remains still unclear but it is possible that it could influence ring opening or coordinates protein dissociation events (Bell and Labib, 2016). After loading is completed, the two Mcm2–7 hexamers tightly associate via their N-terminal domains in a head-to-head configuration, providing a structural basis for the establishment of bidirectional replication forks at the origin (Li et al., 2015b). The number of MCM deposited along the genome is higher than the number of active origins per cell cycle, and the excess MCMs might serve as a back-up in case of replication stress (Raghuraman et al., 2001).

The melting of the DNA to create single strand filaments to be used as templates is a critical step of replication to take place only when cells transition from the G_1 to the S-phase. This mechanism, based on the temporal separation between origin licensing, consisting in the aforementioned steps culminating in the MCMs loading, and their actual firing, provides the basis for ensuring that each ORI is fired only once per cell cycle.

S-phase kininases activate the fork helicases at replication onset

At S-phase transition, two protein kinases, Cyclin-Dependent Kinase (CDK) and Cdc7-Dbf4 (DDK), phosphorylate different targets, resulting in the physical separation of the two Mcm2-7 hexamers, the assembly of the Cdc45-Mcm2-7-GINS (CMG) active replication fork helicases and DNA unwinding (Aparicio et al., 1997; Tercero et al., 2000; Zou and Stillman, 1998) (Figure 10c,d). Both the levels of CDK and of the Dbf4 subunit of DDK are regulated via proteasomal degradation in such a way that their quantity is lower in G_1 and higher in S-phase. Among the targets phosphorylated at this stage, Sld2 and Sld3 play a particularly relevant role (Tanaka et al., 2007; Zegerman and Diffley, 2007). The phosphorylation of the two proteins allows their interaction with Dbp11 that functions as a bridge to recruit the GINS and Pol-ε. The GINS complex (from the Japanese *go-ichi-ni-san* meaning 5-1-2-3, after the four related subunits of the complex Sld5, Psf1, Psf2 and Psf3) is

required both for replication fork establishment at origins and, importantly, also for fork progression (Kamada, 2012; Takayama et al., 2003).

Right after their assembly, the two CMG helicases start unwinding the double helix in a bidirectional manner away from the origin generating ssDNA that is immediately coated by the Replication Protein A (RPA).

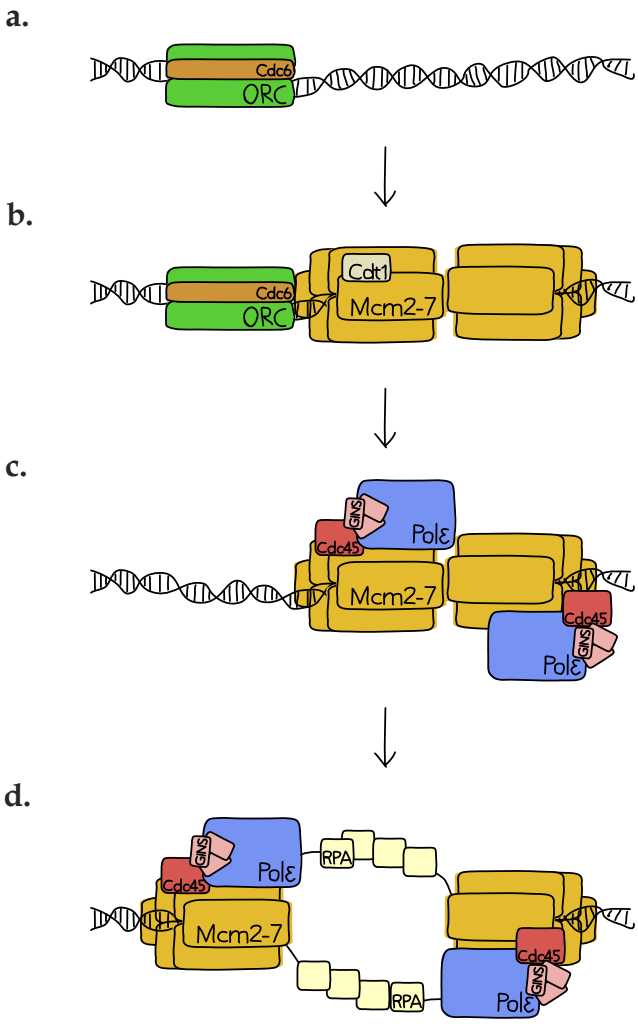


Figure 10. Replisome assembly in eukaryotes.

(a) The ORC complex is loaded at the origin with the co-loader Cdc6. **(b)** Two non-active MCM helicases are loaded asymmetrically in a two-step mechanism. **(c)** At S-phase transition kinases phosphorylate different targets and promote the assembly of CMG active helicases. **(d)** The double helix is unwound in a bidirectional manner and ssDNA is coated by RPA.

II.II Replication elongation

At this stage, everything is settled to start the polymerisation of the new copy of each filament via the action of DNA polymerases.

A "replication timing program" differentiates ORIs in early or late

Before describing the following steps of replication, it is important to notice that not all replicative forks fire simultaneously. Indeed, ORIs can be classified into early or late, according to the moment in which they will promote DNA synthesis (Raghuraman et al., 2001). This differential firing establishes a temporal order, or a "replication timing program", in the synthesis of each DNA segment. The presence of such mechanism is found among all species but is the temporal order of replication to be conserved and not the initiation sites, suggesting that the replication timing is organised by mechanisms that are independent from the ones that specify the origin (Rhind and Gilbert, 2013). Despite this evolutionary generality, the biological significance of replication timing variability remains elusive, but a possibility is that starting replication in different moments during S-phase allows for the activation of back-up ORIs in situations in which forks are stalled or collapsed, whereas if all forks would be activated at the same time the mechanisms preventing re-loading of the MCM helicase in the same S-phase would impede completion of replication (Bell and Labib, 2016).

The timing of an origin is mostly determined by its chromosomal context. For instance, when exchanging the position of the early *ARS1* and that late *ARS501*, they each assumed the timing of the substituted origin rather than maintaining their original timing (Ferguson and Fangman, 1992).

Three DNA polymerases fulfil different roles at the replisome

Once the origin is fired, the catalysis of replication can start; three different kinds of polymerase participate in the process: DNA polymerase α, δ and ε (Pol-α, Pol-δ, Pol-ε), all belonging to the B class of DNA polymerases, but each of them with a specific role.

Pol-α synthesizes RNA primers

Pol-δ and Pol-ε are responsible for replicating the genome enchaining dNTPs complementary to the DNA template in the 5'-3' direction, but none of the two enzymes can start synthesis *de novo* (i.e., without a

primer). The issue is solved by Pol-α, the sole with a primase activity and recruited via interaction with both RPA and the fork helicases for synthesizing ~10 bp RNA primers, the 3'-OH of which is used to start polymerisation by Pol-δ and Pol-ε (Pellegrini, 2012). Indeed, Pol-α lacks an intrinsic 3′ exonuclease activity for proofreading errors and has poor processivity (Perera et al., 2013). Therefore, after synthesis of the RNA primers, DNA replication is carried on by Pol-δ and Pol-ε, better suited for efficient and faithful chain elongation.

The lagging strand is synthesized as smaller intermediary Okazaki fragments

As mentioned above, DNA synthesis is carried out extending RNA primers in the 5′-3′ direction. While the 3′-5′ ssDNA template, named leading strand, can be easily copied as soon as it gets unwound from the double helix, reiterative cycles of synthesis of an RNA primer and its extension are required to duplicate the other filament, hence named lagging strand. The existence of these replication intermediates, long about 100 - 200 bp, was proved in the 1960s by the Japanese molecular biologists Reiji and Tsuneko Okazaki (Okazaki et al., 1968), and hence were named Okazaki fragments.

Table 2. Eukaryotic DNA polymerases.

	Pol-α	**Pol-δ**	**Pol-ε**
Activities	Polymerase	Polymerase 3′-exonuclease	Polymerase 3′-exonuclease
Fidelity	10^{-3}-10^{-4}	10^{-4}-10^{-6}	10^{-5}-10^{-6}
Strand displacement	n.d.	yes	no
Processivity	low	low	high
Interactions	Mcm10, Pol-δ, Ctf4	PCNA, Pol-α	Cdc45, GINS, PCNA, Ctf4

Pol-ε is the leading strand polymerase, while Pol-δ is the lagging strand polymerase

The relative contribution of the Pol-δ and Pol-ε in the synthesis of the leading and lagging strands has been largely studied, yet it is not completely solved. The most widely accepted model, mainly based on the usage of mutator alleles of these two polymerases, proposes Pol-ε to be responsible for the synthesis of the leading strand, while Pol-δ would instead replicate the lagging strand, in collaboration with Pol-α required for the synthesis of RNA primers at the beginning of each Okazaki fragment (Nick McElhinny et al., 2008; Pavlov et al., 2006; Pursell et al., 2007; Shcherbakova and Pavlov, 1996). As mentioned above, Pol-ε is thought to be recruited even before the synthesis of the RNA primer, via interaction with Dbp11, whereas Pol-δ loading may only occur as one of the last steps in replication fork biogenesis, after loading of the Proliferating Cell Nuclear Antigen (PCNA) by RFC (Lee et al., 1991; Tsurimoto and Stillman, 1990). The sliding clamp PCNA acts as a platform to which different accessory proteins bind during replication and integrates multilevel information (Mailand et al., 2013).

Okazaki fragments can be processed via two different pathways

As stated before, the lagging strand is discontinuously replicated as intermediate Okazaki fragments. Each of these fragments must undergo further processing to generate a unique DNA filament lacking gaps as well as RNA primers. Two different pathways exist for the maturation of Okazaki fragments, the 'short flap' and the 'long flap' pathway. The former is believed to be the most frequently used, while the latter is a redundant pathway employed only in particular occasions.

The "short flap" pathway

In the short flap pathway, at the end of each cycle of synthesis of an Okazaki fragment, Pol-δ starts strand displacement synthesis, whereby the newly synthesized DNA displaces the primer of the downstream fragment. This synthesis results in the generation of a 5' flap structure that is subsequently recognised and cleaved by the flap endonuclease Rad27 (Fen1 in human) (Bambara et al., 1997; Gloor et al., 2010; Li et al., 1995). Finally, the remaining gap is immediately ligated by Lig1 to reconstitute the integrity of the filament (Bambara et al., 1997).

The "long flap" pathway

The long flap pathway takes place when, in absence of Rad27, the displacement of the primer proceeds and generates a longer flap, which

at the length of about 22 nt is coated by RPA. The flap is then cleaved by a second endonuclease, Dna2 (Ayyagari et al., 2003; Bae et al., 2001).

Given the incredible dynamism of the process, accessory mechanisms are needed to efficiently coordinate and couple leading and lagging strand synthesis. The trombone model, supported by electron microscopy and single molecule studies, proposes a physical link between the two polymerases on both strands thanks to which coordinated replication DNA takes place by bending the lagging strand back upon itself (Alberts et al., 1983; Chastain et al., 2003; Hamdan et al., 2009). Recent studies have suggested that Ctf4 works as a hub protein containing docking sites for both Pol-α on the lagging strand and GINS and Pol-ε on the leading strand, thereby linking the two machineries (Simon et al., 2014; Villa et al., 2016a).

II.III Replication termination

As discussed in previous paragraphs, each replication origin assembles two diverging forks that travel bidirectionally. Because origins are interspersed along the chromosome, two converging forks meet in regions included between two origins to complete the duplication of the whole molecule. Replication termination is a very delicate and complex, yet poorly understood process that has received little attention during the past years. However, some common characteristics that are shared in bacteria and eukaryotes have been identified. For instance, the two converging forks encounter one another in a region named termination region to complete DNA synthesis. Also, this encounter triggers the disassembly of the replication machineries. In this paragraph, I will discuss the recent advances that have been made on the subject, focusing first on *E. coli*, being the organism for which the process had been best characterised. I will then describe the eukaryotic termination system proposed from evidence coming from the yeast and metazoan models.

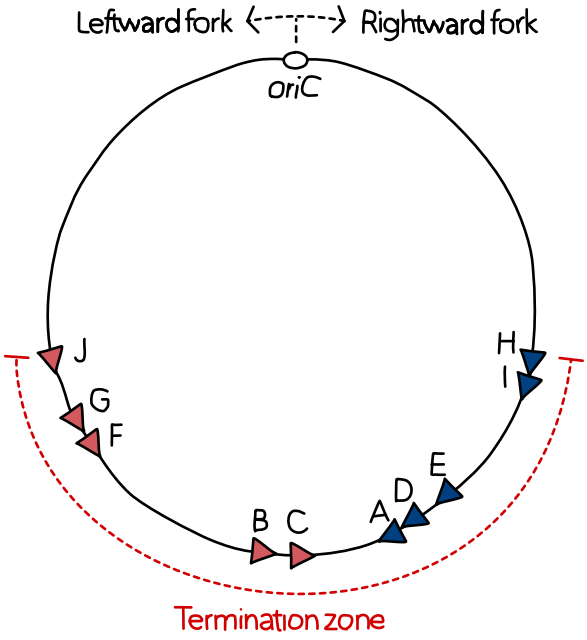


Figure 11. Replication termination on the bacterial chromosome.
Schematic representation of the *E. coli* chromosome, including the *oriC*, and the ten *ter* sites (A–J) depicted as red and blue arrowheads. The termination zone is indicated by a dashed line in red. The *ter* sites are oriented such that the leftward fork can pass the first five *ter* sites that it encounters (red arrowheads), but stalls at the next five sites. Conversely, the rightward fork passes through the *ter* sites marked as blue arrowheads and stalls at the following sites.

A single termination event completes bacterial replication

The single bacterial chromosome contains only one origin implying that a single termination event takes place for each replication cycle. Indeed, the two converging forks meet in a specialised region (Kuempel et al., 1977; Louarn et al., 1977) containing 10 *ter* sites (A-J) (Figure 11), which are binding sites for the DNA replication terminus site-binding protein (Tus), and potent polar replication barriers (Dimude et al., 2016). The *ter* sites are organised in a way that their orientation specifically impacts the progression of one of the two forks: two groups of five sites are placed in tandem and convergently to allow progression of the replication fork traveling from upstream but to impede its further progression after the first five permissive *ter* sites. The subsequence steps are still strongly debated, but the model that has more

experimental evidence proposes that the encounter between the two fork entails the creation of a 3'-flap that is subsequently digested, after which the remaining gap would be filled and ligated to ensure integrity of the molecule (Dewar and Walter, 2017). Beside confining forks to a specific region, the role of the Tus protein remains still unclear, but *tus* mutant have no phenotypes and complete replication normally, thus arguing against an essential requirement of the protein for the completion reaction (Duggin et al., 2008).

In eukaryotes termination zones are dictated by the initiation sites

Unlike the bacterial system, termination regions in eukaryotes lack specific sequences or binding sites for dedicated factors and appears to be mainly determined by the initiation pattern. Briefly, two strong consecutive origins will meet in the midpoint between them regardless of the sequence determinants contained in it (McGuffee et al., 2013). In the most widely accepted model (Figure 12), the two converging forks bypass each other without any clashing until they reach the ssDNA-dsDNA junction of the last Okazaki fragment of the converging fork (Dewar and Walter, 2017). The entering into a dsDNA region is likely to trigger the dismantling of the replication forks, mediated by the E3 ubiquitin ligase SCF which adds K48-linked ubiquitin chains on Mcm7 (Moreno et al., 2014). SCF is tethered to the replisome via the F-box protein Dia2, the substrate-binding component of SCF, via a tetratricopeptide-repeat domain in its N-terminal which binds both Ctf4 and Mrc1 (Mimura et al., 2009). The removal of the CMG makes room for the processing of the final Okazaki fragment, likely by *de novo* recruitment of Pol-δ and by FEN1, thus sealing an intact DNA molecule.

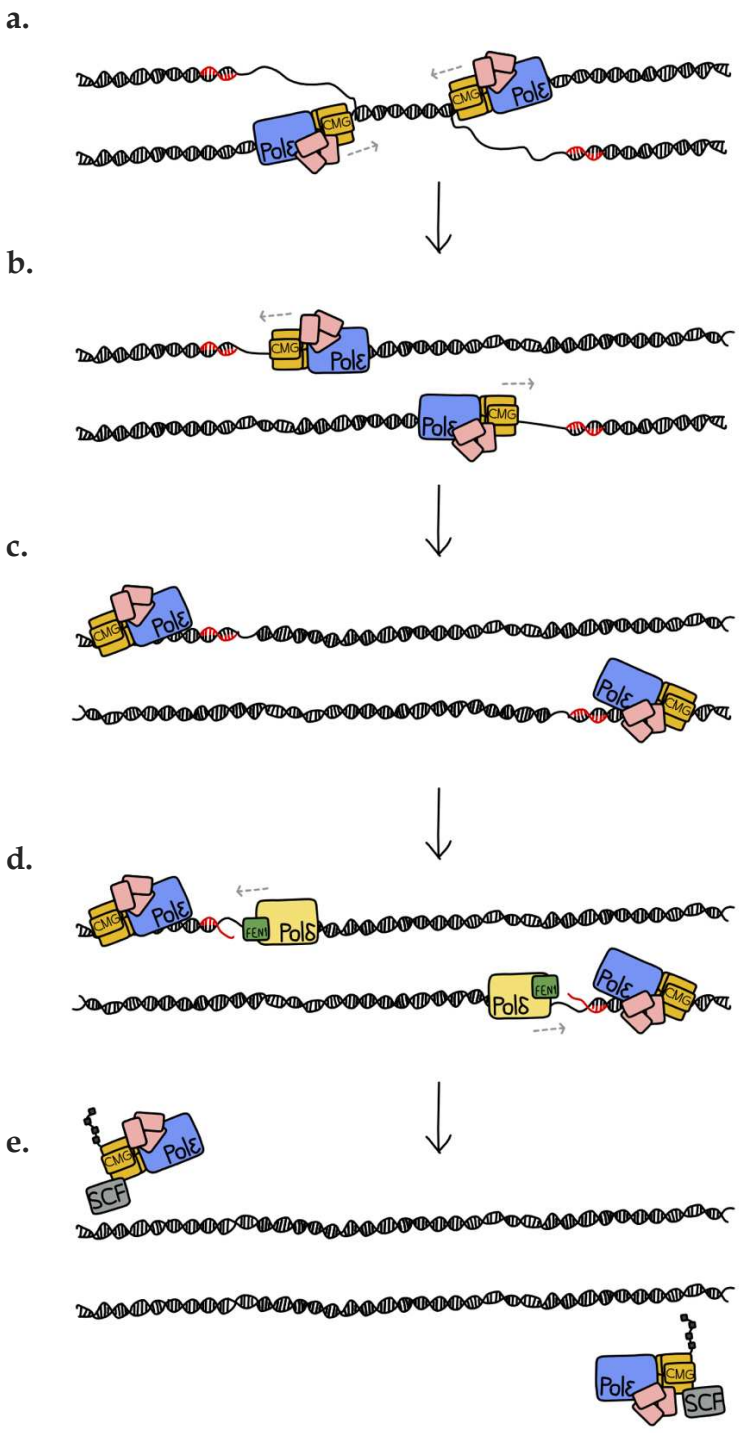


Figure 12. Replication termination in eukaryotes.

(a) After copying most of the replicon, two forks converge into each other. Fork direction is indicated by grey dashed arrow. The red portion of the double helix depicts an RNA primer. **(b)** The two forks bypass each other without any clash. **(c)** The CMG helicases keep translocating until they pass over the ssDNA-dsDNA junction of the downstream Okazaki fragment. **(d)** The final Okazaki fragment is processed by Pol δ and FEN1. **(e)** Once CMG encircles dsDNA, it undergoes polyubiquitylation on its MCM7 subunit by SCF^{DIA2}.

II.IV Topological stress

DNA melting induces rotation of the molecule

When exploring DNA-related process it is inevitable to discuss about the topological properties of the DNA molecule. Indeed, the DNA is subjected to physical constrains each time the double helix must be unwound, during transcription or replication. In the canonical B form, the two complementary filaments intertwine around each other every 10.4 base pairs (Wang, 1979). Any time a bubble of melted DNA is opened along the molecule, the upstream and downstream portion rotate negatively (anticlockwise to the axis of the DNA molecule) and positively (clockwise), respectively (Keszthelyi et al., 2016; Liu and Wang, 1987) (Figure 13). In a short naked linear DNA molecule, the torsion introduced by the melting is easily dispersed by these rotations. However, the long nature of eukaryotic chromosomes as wells as the presence of DNA-bound protein complexes hinder the free rotation of the molecule leading to the introduction of supercoiling.

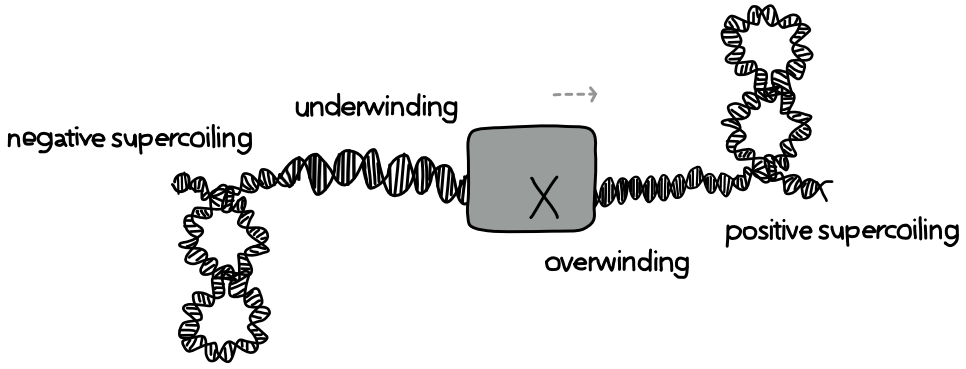


Figure 13. Topological changes induced by transcription or replication.
When a translocating machinery is not allowed to rotate around the DNA axis it introduces overwinding (positive supercoiling) downstream and underwinding (negative supercoiling) upstream its position. The overwinding accumulated ahead prevents strand opening and can ultimately block the movement of the machinery. Underwinding generated behind can promote strand opening and lead to the stabilization of R-loops and other secondary structures.

When molecular machineries such as the RNAP or the replisome travel along the DNA and continuously melt one region after the other, the introduction of more and more supercoiling becomes problematic, up to the point that melting cannot continue if the intertwining is not dispersed. Moreover, during replication the introduction of negative supercoiling behind the fork can lead to the interwinding of daughter molecules and the formation of 'precatenates'. When two forks merge, the precatenates are converted to 'catenates' and the two daughter molecules are physically linked to one another and cannot be separated during chromosome segregation (Dewar and Walter, 2017).

Topoisomerase enzymes can alter DNA topology

Topoisomerases are specialised enzymes that have evolved specifically to bind to the DNA and alter its topology by cleaving one or both the strands to let through either the other strand of the same helix or another double strand, to finally reseal the backbone. Topoisomerases are classified as type I or II, according to whether they introduce a transient single or double DNA break, respectively, but the cleavage always involves the formation of a transient phosphodiester bond between one end of the broken strand and a tyrosine in the active site of the topoisomerase (Nitiss, 1998; Wang, 2002). A further classification is based on structural, mechanistic, and evolutionary considerations and defines type IA, IB, IC, IIA, and IIB enzymes (Wang, 2002).

Bacterial Topo I

The *E. coli* genome encodes for four different topoisomerases: two type IA enzymes, DNA topoisomerases I and III, and two type IIA enzymes, DNA gyrase and DNA topoisomerase IV. The bacterial Topo I was the first topoisomerase to be identified and can only relax negative supercoils (Tse and Wang, 1980; Wang, 1971). It is a 97-kDa protein constituted of an N-terminal domain responsible for cleavage and strand passage and including the active-site tyrosine at position 319; a Zn(II)-binding domain that carries three tetracysteine motifs; a C-terminal domain, rich in basic amino acids and contributing to substrate binding. The N-terminal 67-kDa fragment of *E. coli* Topo I was the first type I topoisomerase crystal to be solved (Lima et al., 1994). The domain structures a "base" and a "lid" around a cavity with a diameter of 28 Å, which can accommodate dsDNA, and with the active-site

tyrosine being placed at the entrance of the cavity. Based on structural evidence, an "enzyme-bridging" model was proposed, according to which the enzyme introduces a nick in the ssDNA and bridges the gap through which the intact strand is passed (Brown and Cozzarelli, 1981; Lima et al., 1994; Tan et al., 2015; Wang, 1996). The clamp then closes around the intact strand, and the cleaved strand is sealed. The enzyme then concludes the cycle by reopening to release the passed strand and closing again. Topo I physically interacts with the β′ subunit of RNAP via its CTD, and thus it localises to transcription units (Cheng et al., 2003). Indeed, it is likely that the enzyme mainly functions to prevent accumulation of negative supercoils generated during transcription that could otherwise disrupt DNA metabolism due to the formation of ssDNA regions and the consequently increase of genotoxic R-loops (Drolet et al., 1995), DNA:RNA hybrids in which an RNA molecule (presumably the nascent transcript) reanneals to its complementary DNA template thus displacing a single strand DNA loop.

Table 3. Topoisomerases.

Topoisomerase		Type	Enzyme structure	Proposed mechanism	Relaxation (-)	Relaxation (+)	Supercoiling (-)	Supercoiling (+)
Topo I	Bacterial	IA	Monomer	Strand passage	yes	no	no	no
	Eukaryotic	IB	Monomer	Controlled rotation	yes	yes	no	no
Topo II		IIA	Homodimer	Strand passage	yes	yes	no	no
Topo III		IA	Monomer	Strand passage	yes	no	no	no
Topo IV		IIA	Homotetramer	Strand passage	yes	yes	yes	no
DNA gyrase		IIA	Heterotetramer	Strand passage	yes	yes	yes	no

Topo III

Topo III is a type IA enzyme with significant homology to Topo I and that relaxes and decatenates DNA (DiGate and Marians, 1989, 1992). Its structure strongly resembles the one of Topo I with the addition of two loops, one of which might be responsible for the decatenation activity (Changela et al., 2007; Li et al., 2000; Mondragón and DiGate, 1999).

Deletion of the gene encoding Topo III is viable, thus suggesting that the enzyme has functions that are redundant with the ones of other topoisomerases (DiGate and Marians, 1989).

DNA gyrase

E. coli DNA gyrase is a type IIA topoisomerase heterotetramer formed from two GyrA and two GyrB subunits, each composed of two principal domains: GyrB contains an N-terminal domain responsible for ATP binding and hydrolysis (Wigley et al., 1991) and a C-terminal domain that allows binding to GyrA and DNA (Chatterji et al., 2000); GyrA consists of an N-terminal domain responsible for DNA cleavage and a C-terminal domain that wraps the DNA (Horowitz and Wang, 1987). DNA gyrase has the unique ability to introduce negative supercoils into covalently closed dsDNA in an ATP-dependent manner (Gellert et al., 1976). It also uses ATP hydrolysis to relax positively supercoiled DNA in a reaction equivalent to the introduction of negative supercoils (Sugino et al., 1978). It has also been shown to be capable of decatenation and unknotting reactions in the presence of ATP (Kreuzer and Cozzarelli, 1980; Liu et al., 1980; Marians, 1987). Moreover, the enzyme can relax negatively supercoiled DNA in an ATP-independent reaction (Gellert et al., 1979). Its function is essential in combination with Muk proteins for the establishment of chromosome condensation during segregation (Sawitzke and Austin, 2000), but it also includes the resolution of positive supercoils arising ahead of both the transcription and replication machineries (Hiasa and Marians, 1996; Kreuzer and Cozzarelli, 1979). Furthermore, the introduction of negative supercoils contributes to the initiation of DNA replication and transcription as underwinding the DNA facilitates melting of the origins and gene promoters (Botchan et al., 1973; Funnell et al., 1986).

Topo IV

Finally, DNA topoisomerase IV is a type IIA enzyme that decatenates replication products, relaxes positive and negative supercoils, and knots and unknots DNA using ATP hydrolysis (Crisona et al., 2000; Deibler et al., 2001; López et al., 2012). In *E. coli* the enzyme is a heterotetramer composed of two copies of two subunits known as ParC and ParE, which are homologous to GyrA and GyrB, respectively (Peng

and Marians, 1993a). Nonetheless, despite their sequence similarities, Gyrase and Topo IV have apparently quite distinct cellular roles. First, Topo IV, unlike gyrase, it is unable to introduce negative supercoils into DNA (Peng and Marians, 1995). Topo IV is also roughly 100 times more efficient at decatenation *in vivo* than is DNA gyrase (Zechiedrich and Cozzarelli, 1995). Topo IV was discovered studying gene deletions leading to DNA partitioning defects, which suggests that Topo IV is involved in decatenation and chromosome segregation (Kato et al., 1990). The idea that this enzyme is the main responsible for decatenation is also supported by an *in vitro* replication system demonstrating that Topo IV is highly efficient at unlinking replicated daughter chromosomes (Peng and Marians, 1993b). However, Gyrase mutants have problems decatenating their chromosomes, entailing that the DNA compaction activity by Gyrase is required for the efficient function of Topo IV.

Eukaryotic Topo I

In both *S. cerevisiae* and *S. pombe* three different topoisomerases have been discovered. Topo I belongs to the IB family, is dispensable for survival, and in contrast to the prokaryotic Topo I, is capable of relaxing both positive and negative supercoils (Champoux and Dulbecco, 1972; Soren et al., 2020). Its mechanism of action is also different, and it implicates the rotation of the broken strand around the intact one, a process known as 'controlled rotation' (Stewart et al., 1998; Stivers et al., 1997). Concerning its role *in vivo*, Topo I probably acts in the unreplicated region between converging forks, and on both the negative and positive supercoils linked to transcription. On this line, Topo I has been shown to colocalize with the transcription machinery, predominantly at the TSS, and to directly interact with the RNAPII CTD, which can also strongly enhance the activity of the enzyme *in vitro* (Baranello et al., 2016; Wu et al., 2010). Topo I is also implicated in the processing of rNMPs in genomic DNA into irreversible single-strand breaks (Cho and Jinks-Robertson, 2017).

Topo II

Topo II is the only essential topoisomerase in budding yeast, and it is a type IIA enzyme assembled as a homodimer that relaxes both positive and negative supercoils and catenates and decatenates DNA in an ATP-

and Mg^{2+}-dependent manner (Goto and Wang, 1982; Liu et al., 1980). Topo II has homology to DNA Gyrase and Topo IV: the N-terminal domain aligns with GyrB and ParE whereas the C-terminal domain aligns with GyrA and ParC (Caron, 1999; Lynn et al., 1986). Topo II is absolutely required to decatenate and prepare chromosomes for segregation and its conditional mutation prevents mitosis (Holm et al., 1985; Uemura et al., 1987). Moreover, Topo II has also been involved in the removal of positive supercoils ahead of the replication fork, even though this role seems to be redundant with Topo I and the relative contribution of the two enzyme remains unclear (Baxter and Diffley, 2008; Le et al., 2019). Finally, Topo II depletion causes stalling of the RNAPII in long genes (> 3 Kb), but complementation with a GyrBA enzyme, which relaxes positive supercoils, rescues the phenotype, suggesting that Topo II, but not Topo I, specifically removes positive supercoils ahead of RNAPII *in vivo* (Joshi et al., 2012).

Topo III

Topo III is a type IA enzyme, and is the less understood topoisomerase in budding yeast. The protein is not essential, but its deletion leads to slow growth, increased mitotic recombination and failure to sporulate because of meiotic recombination defects (Gangloff et al., 1999; Kim and Wang, 1992). The protein is functionally linked to Sgs1, a RecQ helicase, maybe to prevent the formation of lethal intermediates during DNA recombination and, possibly, replication (Gangloff et al., 1994; Suski and Marians, 2008).

Similarly to the yeasts, metazoans carry a Topo I enzyme supporting fork movement. On the other hand, two different isoforms of both Topo II and Topo III are usually found in more complex eukaryotes, in both cases named α and β, with different roles in the cell.

II.V The ribosomal DNA and its replication

The high demand for ribosomes to ensure the synthesis of all proteins in the cell is met by the presence in the genome of several copies of

Ribosomal genes are organised as tandem repeats

ribosomal genes organised in tandem repeats, with the number of those differing among species. In budding yeast, about 150 ribosomal DNA (rDNA) repeats are clustered together on the right arm of chromosome XII (Petes, 1979), even though, akin to other species, only some of these units are actively transcribed. Each 9.1 Kb rDNA repeat contains an *RDN5* gene encoding for the 5S rRNA on the W strand, transcribed by RNAPIII, an *RND37* gene encoding for the 35S rRNA precursor on the C stand, transcribed by RNAPI, and by two non-transcribed spacers (*NTS1* and *NTS2*) (Figure 14).

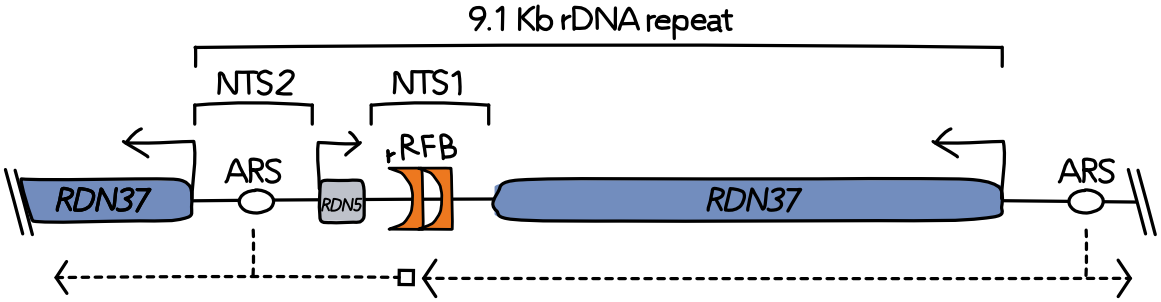


Figure 14. Organisation of the ribosomal unit.
Scheme representing the organisation of the ribosomal unit. Arrows indicate transcription directionality. Dashed arrows indicate replication directionality. The position where the replisome is stalled by the rRFB is depicted as a square.

The ribosomal *locus* is a hard-to-replicate region

The organisation in tandem repeats, combined with the high rates of transcription in this *locus* and the intrinsic difficulty of replicating repetitive sequences, make the rDNA a particularly sensitive region. Moreover, electron microscopy studies on replicating rDNA chromatin showed that both transcription and replication occur simultaneously on the same DNA molecule (Saffer and Miller, 1986), raising the question of how such a crowded environment could be efficiently coordinated to avoid interferences and/or collisions between the two processes. The answer to this question can be at least partially found in the peculiar organisation of the repeat for what concerns the processes of replication and transcription. First, each repeat holds one origin of replication (rARS) placed in the *NTS2* between the *RDN5* and the

RDN35 genes (Linskens and Huberman, 1988; Skryabin et al., 1984), so that the replication forks fired from this origin proceed co-directionally with RNAPI and RNAPIII transcription; second, a strong and polar replication fork barrier (RFB) is located in the *NTS1* (Brewer and Fangman, 1988), immediately after the end of the *RDN5* unit, preventing the replication coming from upstream to enter in the downstream *RDN37* gene, which will be therefore replicated only co-directionally from the next rARS (Brewer et al., 1992; Kobayashi et al., 1992). The co-directionality of transcription and replication has been shown to reduce the risks of negative outcomes from conflicts between the two machineries (García-Muse and Aguilera, 2016). Accordingly, in many bacteria the organisation of the genome is strongly biased towards an orientation of the genes that matches that of the replication forks. In *B. subtilis* the few genes oriented opposite to replication benefit from the mutagenic potential of the head-on collisions to boost adaptation during stress (Lang et al., 2017; Paul et al., 2013).

A polar barrier blocks replication to enter in head-on conflict with transcription

The rRFB is established though the binding to the DNA of the Fork blocking less 1 (Fob1) protein to three different sites (*RFB1*, *RFB2* and *RFB3*) all located in the *NTS1* region (Kobayashi and Horiuchi, 1996). Fob1 is a 65 kDa nucleolar protein containing a zinc finger binding domain, whose integrity is required for DNA binding and induced fork arrest at the rRFB. Fob1 binding is also required for the recombination hotspot activity of the *HOT1* sequence element, which influences the number of rDNA repeats and the formation of extrachromosomal rDNA copies. Despite intense research on the topic, the molecular mechanism by which Fob1 functions to create a polar barrier remains unclear.

The rRFB works through the binding of Fob1

Beside Fob1, three additional factors have been shown to be involved in the rRFB activity: the Topoisomerase I interacting factor 1 (Tof1), the Chromosome segregation in meiosis 3 (Csm3) and the Rrm3 helicase (Ivessa et al., 2003; Mohanty et al., 2006). Rrm3 appears to have a general role in favouring the passage of the replicative forks through proteinaceous barriers, including the rRFB, while Tof1 and Csm3 seem to counteract the action of the helicase. In a recent report, Topo I (Top1)

was also shown to partake in the establishment of the rRFB, where it is recruited via interaction with Tof1 (Shyian et al., 2020).

II.VI Replicative stress and fork stability

Forks are susceptible to collapse and cause DNA damage

During the life of a cell, challenging, unphysiological conditions are not uncommon. Replication is essential for the survival of the next generations and these challenges represent a great risk for the propagation of life (Magdalou et al., 2014; Zeman and Cimprich, 2014). In their course of action, replication forks can encounter DNA lesions, proteinaceous barriers, DNA secondary structures and face conflicts with the transcription machinery. When replication forks enter in contact with one of these elements, they are susceptible to stall and collapse, opening a window for DNA damage and chromosome rearrangements (Lambert et al., 2007; Mirkin and Mirkin, 2007). To minimalize negative outcomes, a surveillance mechanism scans the stability of the forks as well as the integrity of the DNA molecule and operates to solve the damage and resume the replication process as safely as possible (Pardo et al., 2016).

S-phase checkpoints safeguard genome stability

The presence of a corrupted DNA template, which can originate from the action of genotoxic compounds including cellular metabolites, or from ultraviolet (UV) and ionizing radiations, leads to the activation of the DNA damage checkpoint (DDC) (Ciccia and Elledge, 2010). A second signalling pathway is instead dedicated to fork blockage, and it is known as the DNA replication checkpoint (DRC) (Branzei and Foiani, 2007; Tourrière and Pasero, 2007). This last pathway has been extensively studied through the use of drugs inducing fork arrest and collapse, thus increasing the frequency of those events of several orders of magnitude, simplifying their study. Notably, Hydroxyurea (HU) is used to reduce the dNTPs pool by quenching the action of the ribonucleotide reductase and consequently slowing down replication forks (Bianchi et al., 1986; Poli et al., 2012). The alkylating agent Methyl

Methanesulfonate (MMS) is also inducing fork arrest by methylation of the adenines and guanosines in the DNA template, which causes stalling of the replisome and requires a DNA recombination step to restore the fork (Beranek, 1990; Groth et al., 2010).

Mec1 acts as a sensor kinase

Both the DDC and the DRC pathways share Mec1 as sensor kinase and converge on Rad53 as effector kinase (Figure 15), however they differ for the components that mediate the signal from Mec1 to Rad53.

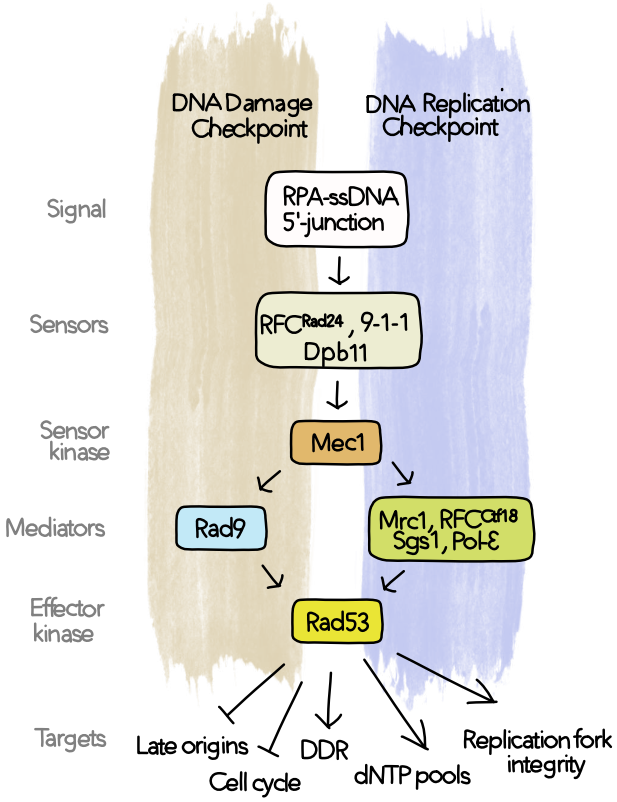


Figure 15. Signalling pathways of replication stress.
Scheme of the two branches of the S-phase checkpoint in *S. cerevisiae*. The DDC and the DRC differ from each other in the mediators used to transduce the signal from the apical Mec1 kinase to the effector kinase Rad53.

The kinases Tel1 and Mec1 are positioned at the top of the DDC signalling pathway. They are both members of the phosphoinositide 3-kinase-related kinases (PIKK) family and are related to mammalian ATM (Ataxia-Telangiectasia Mutated) and ATR (ATM- and Rad3-

related), respectively. The two sensor kinases have redundant roles in terms of initiating the DDC, and they are alternatively recruited depending on the type of lesion but none of the two appears to directly recognise DNA damage. Their loading is rather achieved via the recognition of complexes that physically interact with DNA intermediates generated by damage. Mec1 binds to Ddc2, homolog of the human ATR interacting protein (ATRIP), which recognizes ssDNA bound by RPA (Rouse and Jackson, 2002; Zou and Elledge, 2003), while Tel1 interacts with the C-terminus of Xrs2, component of the evolutionary conserved Mre11-Rad50-Xrs2 (MRX) complex that is able to bind dsDNA ends (Villa et al., 2016b). In S-phase, Mec1 is also recruited to RPA-coated ssDNA at arrested forks to trigger the DRC.

Rad53 is the effector kinase

Other sensor proteins are involved in the activation of the kinases such as Ddc1, Mec3, Rad17, Dpb11, Dna2 and the RFC^{Rad24} complex (Majka and Burgers, 2003; Majka et al., 2006; Navadgi-Patil and Burgers, 2009). Once the sensor kinase is activated, the signal must be amplified to reach all its cellular targets. This is achieved through the effector kinase Rad53, which is transiently recruited to sites of damage and then released to spread the checkpoint response throughout the nucleus (Pellicioli and Foiani, 2005; Sanchez et al., 1996). As already mentioned, the greatest difference between the DDC and the DRC resides at this step, on the transduction of the signal from Mec1 to Rad53. While the former transfers the signal via the mediator protein Rad9, the latter relies on the Mediator of Replication Checkpoint 1 (Mrc1) protein. This is a stable component of the replication forks (Katou et al., 2003), and it has been shown to physically interacts with Pol-ε and Mcm6 (Komata et al., 2009; Lou et al., 2008) and to form, together with Tof1 and Csm3, a Fork Protection Complex (FPC) that fulfils a structural role required for fork progression, but that is distinct from the checkpoint function of Mrc1 (Calzada et al., 2005; Szyjka et al., 2005). When a critical number of replication forks are arrested, Mec1 phosphorylates Mrc1, thus allowing its interaction with the FHA domain of Rad53 to promote its activation (Tanaka and Russell, 2004). Rad53 activation also requires the presence of a modified RFC complex, in which the largest subunit

Mediator proteins differentiate the DDC from the DRC

Rfc1 is replaced by Ctf18, Dcc1 and Ctf8 (Crabbé et al., 2010; García-Rodríguez et al., 2015).

Rad53 targets multiple processes to promote survival

The final steps of the signalling cascade consist in the activation of the actual response aimed at promoting cell survival and that relies on the concerted action of several mechanisms. First, the immediate arrest of the cell cycle to impede entry in mitosis, via the stabilisation of the securin Pds1, which has a key role in sister-chromatid cohesion and in the regulation of spindle elongation (Clarke et al., 2001; Zhou et al., 2016), and the inhibition of the mitotic exit network by Rad53 (Hu et al., 2001; Zhou et al., 2016). Second, the firing of the late origins is repressed to prevent the copy of damaged templates and to preserve backup forks from which to resume replication once the damaged is repaired. This is achieved through the phosphorylation by Rad53 of the two replication initiation factors Sld2 and Dbf4 (Lopez-Mosqueda et al., 2010; Zegerman and Diffley, 2007). Third, the induction of DNA damage response genes (Allen et al., 1994). Fourth, the upregulation of the dNTP pools, through the induction of *RNR* genes expression (Chabes et al., 2003). Fifth, yet unclear mechanisms act to safeguard the replication fork integrity, which is a requisite to allow DNA synthesis resumption, or otherwise face an irreversible collapse. This is supported by several observations including the fact that *rad53* and *mec1* cells exposed to DNA damaging agents fail to restart replication even after drug removal (Morafraile et al., 2015; Tercero and Diffley, 2001), accumulate pathological structures resembling broken forks (Lopes et al., 2001; Sogo et al., 2002) and lose components from replisomes (Cobb et al., 2003, 2005; Katou et al., 2003).

III. R-Loops and Transcription-Replication Conflicts

The act of transcription itself carries intrinsic risks due to different aspects of the nature of the process. Among them, during the past year, increasing attention has been dedicated to R-loops and to Transcription-Replication Conflicts (TRCs). Both phenomena are susceptible of causing damage in the genome, and therefore, commonly to other processes that can lead to genome instability, intricated mechanisms of control have been revealed to exist. Moreover, some connections exist between R-loops and TRCs. Along this chapter, I will focus on the nature and the metabolism of R-loops, on the current knowledge about TRCs, and on some of the proteins involved in either one or both the processes, the function of which has been also central for this study. Finally, I will illustrate the currently available techniques for the detection of R-loops, in order to better appreciate the last part of my work which aimed to devise a novel R-loop detection method.

III.I R-loops metabolism

DNA:RNA hybrids can form also outside the RNAPII inner channel

As described in Chapter I, when transcription takes place, the dsDNA is unwound in order to use the ssDNA as a template for the polymerisation of the nascent RNA which occurs in the inner channel of the RNAP thus forming a short DNA:RNA hybrid. As the transcription machinery proceeds, the DNA duplex reanneals upstream of the polymerase, while the RNA coming out from the exit channel of RNAPII is coated by RNA binding proteins. Nevertheless, in some cases, the nascent transcript can form an DNA:RNA hybrid outside the RNAP inner channel and that can extend for several nucleotides, by hybridising to the DNA template and extruding a ssDNA portion, thus resulting in a structure known as R-loop.

The "extended DNA:RNA hybrid" model

It is yet somehow unclear how R-loops are formed, and two models have been proposed so far. The first one, known as the ''extended DNA:RNA hybrid'' model, proposes the DNA:RNA hybrid to be an extension of the short 8 bp DNA:RNA hybrid formed within the transcription bubble during RNAP transcription. Nevertheless, the crystallographic structure of RNAPII demonstrates that the DNA and RNA molecules exit through different channels (Westover et al., 2004), hence strongly arguing against this model. The second one, called the ''thread back model'', proposes a more plausible mechanism suggesting that the DNA:RNA hybrid forms by threading back the RNA to the template before the two DNA strands in the transcription bubble reanneal.

The "thread back" model

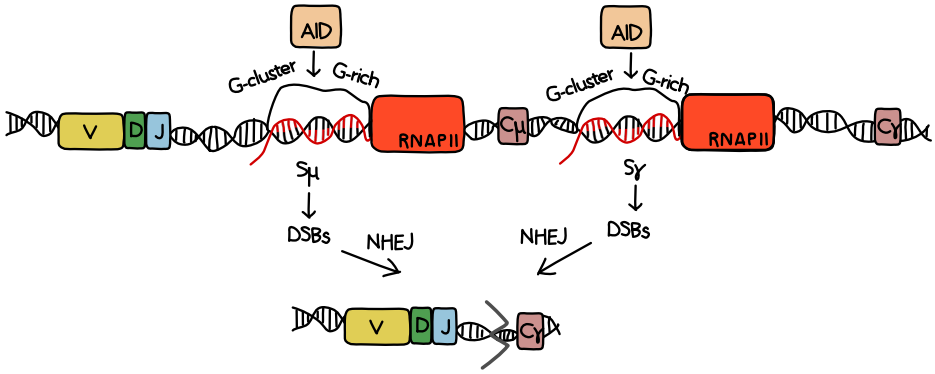
Nucleotide composition influences the formation of R-loops

Not all transcribed regions form R-loops to the same extent. Several studies have pointed out a preference for GC-rich regions (Ginno et al., 2012, 2013; Li and Manley, 2005; Reaban et al., 1994), even though also AT-rich regions appear to be prone to generate hybrids (Wahba et al., 2016). In particular, R-loops are formed preferentially when the non-template strand is G-rich which could be explained by the increased thermodynamic stability of a G-rich RNA strand bound to the C-rich DNA strand (Sugimoto et al., 1995). This arrangement suggests that stable secondary structures such as G4 quadruplex may form on the non-template strand coincidentally with R-loops. Consistently, the presence of a G4 within a non-template G-rich strand has been shown to favour the generation of R-loops (Duquette et al., 2004, 2007). Additionally, a number of specific DNA regions have shown to display increased tendency to form R-loops both in yeast and human cell lines, such as the rDNA, the mitochondrial genome and repetitive regions among which yeast transposons, human L1 LINEs, and telomeres (Balk et al., 2014; Chan et al., 2014; El Hage et al., 2014; Ginno et al., 2012; Nadel et al., 2015; Wahba et al., 2016).

DNA:RNA hybrids adopt a 'heteromerous' conformation that is intermediate between the B form of dsDNA and the A form of dsRNA and are very stable due to their oligomeric length, the content of deoxypyrimidines/deoxypurines, and the A·T/U proportion (Fedoroff

OYu et al., 1993; Salazar et al., 1993; Shaw and Arya, 2008). Consequently, once formed, their removal can be an energy consuming process.

a.



b.

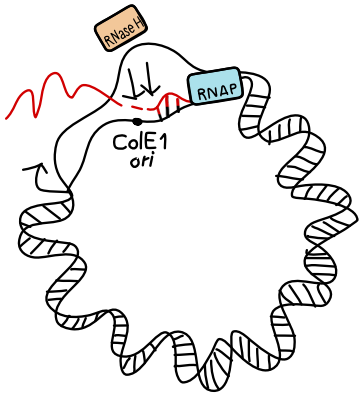


Figure 16. The roles of R-loops in class-switch recombination (CSR) and replication.
(a) CSR consists in the recombination event between two switch (S) regions. R-loops are formed during RNAPII transcription, and the G-rich non-template strand is targeted by AID, thus promoting the formation of DSBs and a NHEJ to repair the damage. **(b)** R-loops in ColE1-type plasmids replication.

Despite having been shown to form across the genomes of bacteria, yeast, and more complex eukaryotes, for many years R-loop formation was thought to be a rather rare event, and merely a by-product of transcription. However, increasing evidence collected relatively

recently has revealed a much more articulated picture. A controversial characteristic of these hybrids resides in their impact that can be easily seen as a double-edged sword. From one side, R-loops have been shown to be involved in physiological processes, whereby they exert a rather positive role, such as DNA recombination, replication, and transcription. The best characterized example of the first kind is the Ig class-switch recombination (CSR) (Figure 16a). CSR takes place in specific repetitive switch (S) regions, where formation of an DNA:RNA hybrid displaces a G-rich ssDNA region that is targeted by the B cell-specific cytidine deaminase AID to generate double strand breaks (DSBs) and stimulate recombination (Huang et al., 2007; Yu et al., 2003).

R-loops play a role in CSR and in replication

The replication of the bacteriophage T4, of the *E. coli* ColE1 plasmid (Figure 16b), and of the mitochondrial DNA provides instead examples of R-loop involvement in DNA replication. In all these cases, an DNA:RNA hybrid produced from an RNAP is used to generate a 3'-end that is then extended by a DNA polymerase (Itoh and Tomizawa, 1980; Kreuzer and Brister, 2010).

Experimental evidence links R-loops to transcription termination

The impacts of R-loops on transcription are more variegated as enlightened by several reports unveiling very different aspects that can be influenced by the formation of DNA:RNA hybrids. It was shown that R-loops forming in the G-rich 5'-UTR regions located downstream of CpG-non-methylated promoters inhibit *de novo* DNA methyltransferases thus influencing the epigenetic state of these *loci* (Ginno et al., 2012). In human cells, the formation of DNA:RNA hybrids at the termination site of genes allows the recruitment of the helicase Senataxin (SETX), the human homologue of Sen1, which subsequently releases the nascent RNA and thus allows its Xrn2-mediated degradation and transcription termination (Skourti-Stathaki et al., 2011). A link between R-loops and transcription termination is supported also from genome-wide analysis revealing that G-rich sequences are very commonly found immediately downstream from the poly(A) signal in mammalian genes (Salisbury et al., 2006) and that promoters and 3' regions of genes are enriched in G4-forming sequences (Huppert et al., 2008). Interestingly, the presence of G4s is

favoured between consecutive genes, further suggesting an involvement in transcription termination. Given the aforementioned potential of G-rich sequences to form R-loops, it can be envisioned that R-loops are intrinsic elements of termination pause sites. On this line, G-rich sequences, on which stable R-loops are formed, stop T7 RNAP transcription (Belotserkovskii et al., 2010). In bacteria, mutants with defects in Rho-dependent termination display R-loops accumulation, as these strains require RNase H activity for survival (Hong et al., 1995), which may provide an alternative route for processing and degradation of mRNAs (Anupama et al., 2019).

R-loops have also been involved in the silencing of long non-coding RNAs (lncRNAs). Examples cases are the lncRNA COOLAIR in *Arabidopsis thaliana* and the Ube3a antisense transcript in human cells. Upon cold exposure, COOLAIR inhibits the expression of a key repressor of flower development, the FLC gene. The silencing of COOLAIR itself is controlled by the formation of R-loops over its promoter, where a ssDNA-binding homeodomain protein, AtNDX, binds and stabilizes these R-loops, ultimately leading to COOLAIR transcriptional repression (Sun et al., 2013). Similarly, R-loops inhibit the expression of the Ube3a antisense transcript, associated with the Prader-Willi syndrome (Powell et al., 2013).

R-loops can have genotoxic consequences and RNA biogenesis prevents their formation

From the other side of the medal, R-loops have been shown to also have a strong genotoxic potential and have been often described as threats to genome stability. The first evidence comes from studies on yeast mutants of THO, a complex involved in transcription and RNA export. These strains were shown to accumulate R-loops, transcription-associated hyper-recombination phenotypes and to present elevated chromosome and plasmid loss; all phenotypes that could be rescued by overexpression of RNase H1, thus implicating R-loops as the source of DNA damage (Huertas and Aguilera, 2003). On the same line, mutants of different subunits of the cleavage and polyadenylation factors (such as *RNA14*, *RNA15* and *PCF11*), or affected in RNA surveillance (such as *RRP6*), or in transcription termination (such as *SEN1*) all display increased R-loop levels and genomic instability that can be overall

rescued by RNase H overexpression (Luna et al., 2005; Mischo et al., 2011). Consistently, a screen to identify mechanisms causing increased chromosome instability (CIN) in budding yeast pinpointed mRNA biogenesis factors and also in this case the CIN was suppressed by RNase H1 overexpression (Stirling et al., 2012). A genome-wide siRNA screen in human cells aimed at identifying genes whose silencing would lead to H2AX phosphorylation as well yielded a number of RNA-processing factors, and, similarly to other studies, this was shown to be a RNase H1-dependent effect (Paulsen et al., 2009). Because of these clear links between R-loops levels and genomic instability, hybrids formation is tightly monitored at the cellular level. Consistent lines of evidence support the notion of a link between RNA metabolism and R-loop control. It has been argued that efficient mRNA packaging into mRNPs hampers hybridisation of the RNA molecule to the DNA (Figure 17b) and introns and/or splicing factors likewise reduce R-loops formation (Bonnet et al., 2017; Li and Manley, 2005). The DNA topology status also influences R-loops formation. The accumulation of negative supercoiling eases DNA melting, thus lowering the energetic barrier that must be overcome to form an R-loop. It is therefore not very surprising that Top1, which relaxes negative supercoils, is implicated in R-loops prevention in several organisms (Drolet et al., 1995; El Hage et al., 2010; Tuduri et al., 2009).

R-loops are eliminated by their unwinding or digestion

Beside preventing their formation, cells also employ a series of mechanisms to actively eliminate R-loops once they arise. The factors that have been shown to play preponderant role in this sense are the RNases H enzymes, which degrade the RNA moiety of these structures, and several helicases, including Sen1, which have been proposed to unwind heteroduplexes (Figure 17a). Sen1 and RNases H will be described in detail in the following paragraphs.

R-loops and DNA damage

Even though R-loops can cause DNA damage, the underlying mechanisms are not fully elucidated. One argument is that the unpaired ssDNA is more susceptible to spontaneous DNA damage such as the deamination of dC to dU, which induces DSBs and mitotic recombination (Aguilera, 2002; Aguilera and García-Muse, 2012; Li and

Manley, 2006). Another possibility is that the R-loops provide an appropriate substrate targeted by the action of specific mutagenic enzymes such as AID (Chaudhuri and Alt, 2004; Gómez-González and Aguilera, 2007; Petersen-Mahrt et al., 2002). Nevertheless, AID is only expressed in activated B cells and in chicken DT40 cells, thus raising the question of how R-loops would be targeted in other cell types and/or organisms. An additional argument is that R-loops induce genomic instability by interfering with DNA replication, an hypothesis that is supported by the fact that collisions between the replisome and blocked RNAPIIs lead to increased recombination and DNA breaks (Gottipati et al., 2008; Prado and Aguilera, 2005). DNA lesions in the ssDNA of the R-loop or the DNA:RNA hybrids themselves arising upstream of RNAPII may somehow restrict transcription, which in turn may block replication forks. In a very recent study, an *in vitro* reconstituted system was used to show that DNA:RNA hybrids and G4s interfere with replisome progression by inducing fork stalling (Kumar et al., 2021).

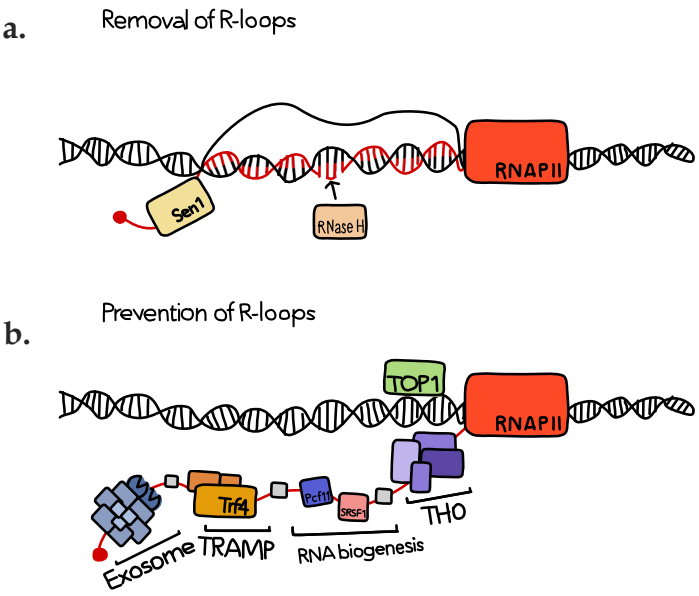


Figure 17. Cellular factors influencing R-loops levels.
(a) R-loop removal can be promoted by its degradation (RNases H) or by unwinding (helicases). **(b)** R-loops are prevented by reducing negative supercoiling, which relaxes DNA, or by disfavouring RNA hybridisation through its processing.

A puzzling question is how R-loops can be both beneficial and a threat to genome stability, and more specifically how the physiological regulatory R-loops do not drive DNA damage and seem somehow to be protected from induction of DSBs. Despite the absence of empirical observations to answer this question, it appears reasonable to think that regulatory and genotoxic R-loops must differ between each other for at least one or more features, such as their size, number, persistency, or location. A recent study in budding yeast showed that only a subset of very persistent R-loops leads to irreparable DNA damage and cell death (Costantino and Koshland, 2018). Another intriguing observation is that genotoxic R-loops stimulate H3 Ser10 phosphorylation, a mark associated to chromatin compaction, and which might induce replication fork stalling, transcription-replication conflicts and DSBs (Castellano-Pozo et al., 2013). Atomic Force Microscopy (AFM) studies showed that different genes with similar tendency to form R-loops *in vitro* resulted in a different architectural organization of the ssDNA loop, likely because of their specific sequences, and this could provide another mean to differentiate R-loops (Carrasco-Salas et al., 2019).

III.II Transcription-Replication Conflicts

DNA and RNA polymerases both function by copying a DNA template, and therefore compete for the same substrate, thus raising the question if the two machineries can hinder each other, and if so, how such interferences are prevented and/or solved. An intuitive solution would be to confine RNAPs activity outside S-phase, hence preventing any encountering with the replication forks. However, several factors are specifically required during S-phase, and the temporal activation of a subset of genes during this phase allows the establishment of layers of control that are important for survival. Considering the large number of RNA polymerases on the genome, particularly so in a context of

widespread non-coding transcription, Transcription-Replication Conflicts (TRCs) are expected to occur frequently.

Transcription can interfere with replication

The question began to be addressed during the 80's using an *in vitro* reconstituted system that showed that the bacteriophage T4 replication complexes were blocked by the presence of *E. coli* RNAP bound to dsDNA (Bedinger et al., 1983; Jongeneel et al., 1984). If transcription was allowed in the same direction as replication, replication forks appeared to adapt to the moving RNAPs at the slower rate of transcription. Addition of the T4 bacteriophage *Dda* helicase to the reaction removed the block created by the RNAP restoring normal DNA replication. Moreover, other indirect observations started to point to the existence of conflicts. For instance, it was noted that impeding transcription initiation would hasten DNA synthesis, while, on the contrary, perturbing transcription elongation would result in the opposite effect (Pato, 1975). Additional clues came from the study of genome organisation, when it was noted that the highly transcribed ribosomal genes are co-directionally oriented with replication and that a similar feature is shared by many bacterial genomes (Ellwood and Nomura, 1982; Nomura et al., 1977). In *E. coli* the orientation of highly expressed genes is strongly biased in the same orientation of replication forks, while this is not the case for poorly expressed genes (Brewer, 1988). According to their respective orientation, TRCs can be either co-directional (CD), when the two machineries proceed in the same direction, or head-on (HO), if the directionality is opposite (Figure 18). In prokaryotes, the speed of a replication forks is about 10 times higher than that of a RNAP (Helmrich et al., 2013), so CD encounters can take place when the replisome is positioned upstream of the RNAP, while when RNAP proceeds behind the fork the replisome must be blocked or its speed decreased for a conflict to occur. In a CD conflict, it can be envisioned that the fork would either slows down until the end of the transcription unit when physiological termination of RNAP remove the obstacle, or that it would be able to remove the RNAP from the DNA template. Importantly, the speed of transcription and replication in eukaryotes is rather similar (Helmrich et al., 2013), which decreases the chance of CD conflict. On the contrary, HO conflicts necessarily require

Conflicts can be head-on or co-directional

the dismantling of the EC. Moreover, as already discussed (see Introduction, § II.IV), progression of the fork and the EC introduces positive supercoils ahead of the machineries, which, in a HO conformation, sum up potentially creating strong topological stress that might even impede further unwinding of the double helix. On the contrary, in a CD conformation the positive supercoiling introduced by one of the machineries is cancelled by the negative supercoils introduced by the other.

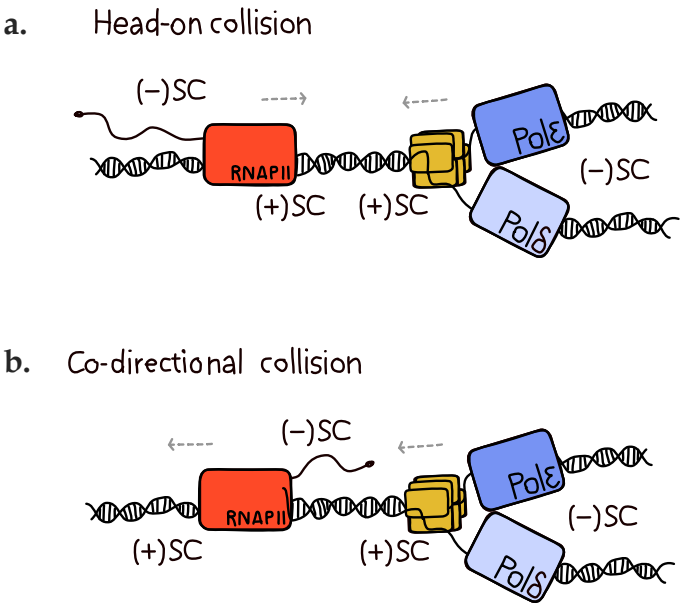


Figure 18. The different configurations of Transcription-Replication Conflicts.
(a) In head-on collisions the two machineries converge, and both introduce positive supercoiling between them. **(b)** In co-directional collisions the positive supercoiling introduced by one of the machineries is cancelled by the negative supercoiling introduced by the other. Movement directionality is indicated by grey dashed arrows.

HO conflicts are more detrimental

Mounting evidence supports the notion that HO conflicts are more detrimental, which is consistent for instance with the already mentioned genome organisation that favours co-directionality for highly transcribed genes (Srivatsan et al., 2010; Wang et al., 2007). A

seminal study employed electron microscopy to determine the directional effect of rDNA transcription on fork progression by inverting the *rrnB* ribosomal operon in the *E. coli* chromosome, thus conferring an artificial HO conformation and revealing that this provokes a reduction in the speed of the forks (French, 1992).

The rDNA and the tRNA genes are RFP sites

At a first glance, Replication Fork Pausing (RFP) in eukaryotes was observed both at tRNAs and in the rDNA (Brewer and Fangman, 1988; Deshpande and Newlon, 1996; Kobayashi and Horiuchi, 1996). In the first case, RFP was shown to occur only when RNAPIII transcription and replication are in a HO conformation, and dependently on an active promoter (Deshpande and Newlon, 1996; Ivessa et al., 2003). However, a recent report revealed that the RNAPIII-transcription factor TFIIIB, and not active transcription, is the main determinant of the pausing of the replisome (Yeung and Smith, 2020). It remains unclear why RFP occurs only in a HO conformation. In the second case, TRCs at the rDNA *locus* in *S. cerevisiae* were detected in a *fob1Δ* strain (see Introduction, § II.V), when the RFB is not active, leading to a reduction of the number of rDNA gene repeats (Kobayashi and Horiuchi, 1996).

Also RNAPII genes can induce RFP

RNAPII transcription units were also shown to be impediments to fork progression. In 2005, Prado & Aguilera employed plasmid-borne recombination constructs in *S. cerevisiae* to show that head-on RNAPII transcription impairs replication fork progression, generating RFP and a significant increase in recombination (Prado and Aguilera, 2005). The Rrm3 helicase was shown to facilitate replication through the transcription-dependent RFP sites and to reduce recombination levels. In ChIP experiments DNAP occupancy was found to be significantly increased in highly transcribed regions, which was explained by decreased replisome speed in these regions (Azvolinsky et al., 2009).

An intriguing question is how the different states of RNAPII can affect replication. A study from 2006 by Mirkin and co-workers engineered a strong bacterial promoter to stabilise the transcription initiation complex and showed by 2D gel electrophoresis that this leads to a HO pausing of the replisome (Mirkin et al., 2006). Surprisingly, the transcription termination site also unexpectedly proved to be a fork

pausing site, but in this case in a CD manner. Thus, the transcription initiation and termination sites create polar barriers for replication, which the authors proposed to function as "punctuation marks" creating extra room for the repair or gene conversion machineries to clear the coding areas off newly acquired mutations.

CD conflicts also appear to induce fork stalling, although to a lower extent (Helmrich et al., 2013).

R-loops and non-canonical B-form are linked to TRCs

It is important to notice that beside RNAPII per se, also by-products of transcription can hinder replisome progression. In particular, the non-B form that particularly repetitive DNA sequences can assume, such as hairpins, triplex DNA or G4, have been shown to be obstacles for the replicative forks (Zhao et al., 2010). The formation of these abnormal structures is facilitated by ssDNA exposure, and transcription, which relaxes the DNA introducing negative supercoils behind the RNAPII might facilitate the process. Formation of co-transcriptional R-loops has also been correlated to fork stalling by several reports, in the yeast and human models and it has been proposed that these hybrids cause replication fork stalling at telomeres, rDNA regions, CpG islands and at specific RNAPII-transcribed genes (see Introduction, § III.I). Hamperl and co-workers (Hamperl et al., 2017) provided proof that TRCs themselves can influence R-loop formation or prevention according to the relative orientation of the transcription and replication machineries. HO conflicts were shown to favour hybrids accumulation at the site of the conflict, whereas CD conflicts to reduce R-loops levels.

Several mechanisms, such as polar barriers, are in place to prevent from TRCs

During the last decade many groups have been doing strong efforts to characterise the mechanisms that are in place to prevent and or resolve TRCs and describe the whole panorama of the actors involved. I have already mentioned that genome organisation plays a role in the prevention of the more dramatic head-on conflicts. However, especially in eukaryotes, but not exclusively, a clean-cut genic organisation that allows co-directionality with replication does not appear to exist, beside for the rDNA *locus*. In this sense, polar replication fork barriers are some of the strategies employed to avoid TRCs, with Fob1 at the rRFB being the best characterised example (see Introduction, § II.V).

Factors affecting RNAPII processivity also influence TRCs

A second mechanism employed to avoid TRCs relies on the transcription apparatus itself. Disfunction or depletion of transcription elongation factors or defects in the recovery of backtracked polymerases result in diminished fork progression (Dutta et al., 2011; Felipe-Abrio et al., 2015; Tehranchi et al., 2010). Similarly, the use of a modified promoter that induces a permanent arrest of the RNAP leads to DSBs in a replication-dependent manner (Dutta et al., 2011; Pomerantz and O'Donnell, 2010). In general, it is possible to argue that these factors do not have a direct role in TRCs, however they affect the stability of the EC and consequently they favour or disfavour its eviction from the DNA. It is reasonable to think that more stable ECs act as better barriers. A notion that is also supported by the observation that *rpb1-1* cells, which shows a tighter binding of RNAPII to the DNA as observed by ChIP, exhibit replication impairment (Felipe-Abrio et al., 2015).

Several helicases play a role in the control of TRCs

Helicases also have been shown to play a role in RFP from bacteria to mammals. In *E. coli* and *B. subtilis*, auxiliary helicases, such as Rep and UvrD in the former and their homolog PcrA in the latter, promote replication through transcribed genes, even though probably via distinct mechanisms (Baharoglu et al., 2010; Merrikh et al., 2015). Rep and PcrA appear to act as fork-specific motor proteins helping replication across transcribed DNA templates, whereas UvrD is likely interacting with arrested RNAPs via the Transcription Coupled Repair pathway. In *S. cerevisiae* the Pif1 family of helicases counts two members: Pif1 and Rrm3, both involved in resolving RFP (Ivessa et al., 2000, 2003; Paeschke et al., 2011). The first has been shown to facilitate fork progression through proteinaceous blocks and G4 structures, while the removal of the second leads to increased RFP at tRNAs and at the rRFB, where this results in breakage and accumulation of excised rDNA circles. Interestingly, Rrm3 also associates with highly transcribed genes in mutants that accumulate R-loops and RNase H overexpression significantly reduced the interaction of the helicase with these regions (Santos-Pereira et al., 2013). The fission yeast homolog of Pif1, Pfh1 is also enriched at highly transcribed genes, and its absence provokes replication fork stalling (Sabouri et al., 2012). In

human cells, depletion of the RECQL5, helicase, which interacts with the replisome component PCNA, induces DSBs and chromosome rearrangements in transcribed genes and at Common Fragile Sites (CFS) (Hu et al., 2009; Li et al., 2011, 2015a; Saponaro et al., 2014). CFS are genomic regions characterized by constrictions or gaps in metaphase chromosomes following replication stress and often enriched in sequences that can stall DNA replication (Glover et al., 1984). Underlying causes of the fragility at these sites have been proposed to be the scarcity of replication origins and the inefficient activation of replication (Letessier et al., 2011; Ozeri-Galai et al., 2011), as well as the concomitant occurrence of transcription and replication (Helmrich et al., 2011). Notably, RECQL5 also binds to RNAPII, and several *in vitro* and *in vivo* studies suggest that it has a negative effect on transcription elongation rate (Popuri et al., 2013), and its depletion results in increased stalling and/or arrest (Saponaro et al., 2014), hence hinting that its function in TRCs might be related to the modulation of transcription.

Chromatin remodelling and Dicer prevents from TRCs

Chromatin remodelling is likely playing a role in preventing TRCs, even though its contribution was addressed less deeply. The best characterised example concerns the histone chaperone FACT. Yeast and human cells lacking FACT complex activity display higher levels of TRCs, fork progression impairment and genomic instability in a transcription-dependent fashion (Abe et al., 2011; Foltman et al., 2013). In *S. pombe* the RNAi machinery silences the pericentromeric regions via the recruitment of chromatin modifiers by siRNA, and it has been proposed that it releases RNAPII in these regions to prevent collisions (Zaratiegui et al., 2011). Further investigations on the role of the RNAi factor Dicer in TRCs has shown that removal of RNAPII might not restricted to pericentromeric regions but also takes place at highly transcribed genes, rDNA and tRNA genes (Castel et al., 2014).

Temporal and spatial separation minimises TRCs in metazoans

In human cells undergoing replication, microscopy studies of the human rDNA cluster have shown that the DNA replication and transcription machineries occupy distinct nuclear territories presumably to limit mutual interference (Smirnov et al., 2014; Wei et

al., 1998). At the genome-wide scale, a coordination exists both at the temporal and spatial level aimed at minimizing TRCs in metazoans: a nascent RNA capture assay revealed an anti-correlation between timing of replication and transcription (Meryet-Figuiere et al., 2014), and a certain degree of co-orientation of replication with transcription was observed, due to preferential (but not exclusive) initiation of DNA replication from upstream of active and highly transcribed genes (Petryk et al., 2016).

III.III Sen1 and genome stability

Sen1 mutants show increased TAR levels

As already described before, Sen1 is a component of the NNS complex and its deletion leads to lethality at least partially linked to widespread transcription termination defects (see Introduction, § I.III.II and Results, Chapter I). Similar transcription termination defects occur in the absence or mutation of any other subunit of the NNS complex (Carroll et al., 2007). However, in 2011 Mischo et al. reported some phenotypes that were specific to *sen1-1*, a thermosensitive allele of *SEN1* bearing a point mutation in the helicase domain (G1747D) (DeMarini et al., 1992), and that were not observed for *ts* alleles of *nrd1* and *nab3*, thus providing evidence for a role of Sen1 beside the NNS complex (Mischo et al., 2011). The authors employed reporter cassettes specifically designed to assess the levels of Transcription-Associated Recombination (TAR). These reporters contain a selection marker disrupted by an insertion flanked by homology regions, and under the control of a strong inducible promoter. Upon induction of transcription, the presence of DNA damage activates the DNA repair response, which leads to the restoration of the coding frame of the selection marker via recombination between the two homology regions flanking the insertion. When performing this assay in NNS mutants, *sen1-1* showed a frequency of TAR events markedly higher than *nrd1* and *nab3* mutants (Mischo et al., 2011). Moreover, the increased frequency of TAR events was sensitive to *in vivo* Rnh201 overexpression, thus leading the

Increased TAR levels in *sen1-1* are linked to R-loops accumulation

authors to conclude that the leading cause of DNA damage in the reporter cassette was the accumulation of genotoxic DNA:RNA hybrids and that Sen1 prevents from the accumulation of these aberrant heteroduplexes during transcription (Mischo et al., 2011). Accordingly, higher levels of S9.6 signal were shown to accumulate in those reporters in *sen1-1* cells and this allele was found to be lethal when combined with deletions of DNA repair genes such as components of the MRX complex, Rad52 and the Sgs1 helicase (Mischo et al., 2011).

Sen1 co-localises with sites of BrdU incorporation

In 2012 Alzu and co-workers provided further evidence for a role of Sen1 in genome stability by showing by ChIP that Sen1 colocalizes with the position of the replisome, inferred by bromodeoxyuridine (BrdU) incorporation, while this was not the case for the NNS subunit Nrd1 (Alzu et al., 2012). Moreover, *sen1-1* cells incubated at non-permissive temperature exhibited the accumulation of aberrant replication forks with concomitant presence of DNA:RNA hybrids at sites of active transcription oriented head-on with the replication machinery (Alzu et al., 2012). Depletion of the MRX or of the RFC complexes abolished the accumulation of *sen1*-dependent aberrant forks and R-loops at sites of conflicts, suggesting that those components have a role in preventing the premature processing of these forks via other pathways, and most likely from the action of the nuclease Exo1, which if conjointly depleted, restored the presence of the defected forks (Brambati et al., 2018).

In a recent report, Sen1 was shown to counteract the formation of DNA:RNA hybrids at HO-induced DSBs leading to an aberrant increased resection and triggering a KU-dependent mutagenic NHEJ (Rawal et al., 2020).

Also SETX plays a role in genome stability

The contribution of Sen1 to genome stability appears to be evolutionary conserved. Indeed, disruption of the mice homologue of Sen1 leads to defects in spermatogenesis due to failure in meiotic recombination (Becherel et al., 2013; Yeo et al., 2015). A wide number of somewhat disconnected results have also pointed towards a role for the human homologue of Sen1, Senataxin (SETX), in genome stability. For example, Sen1 was shown to bind the tumour suppressor breast cancer susceptibility gene 1 (BRCA1), widely involved in genome stability,

and this interaction is required to recruit SETX to a R-loops forming regions, especially at the termination sites of active mammalian genes, presumably to prevent DNA damage (Hatchi et al., 2015). A similar function was as well proposed to be played via interaction with SMN (Zhao et al., 2016). Accordingly, absence of SETX leads to the formation of ssDNA breaks and γH2AX *foci* (Hatchi et al., 2015; Roda et al., 2014; Sollier et al., 2014). Similarly to Sen1, SETX was shown to be recruited at induced DSBs in active *loci* presumably to unwind DNA:RNA hybrids (Cohen et al., 2018). Importantly, mutations of SETX have been linked to the development of Amyotrophic Lateral Sclerosis of type 4 (ALS4) and to Ataxia-ocular apraxia 2 (AOA2) (Chen et al., 2004; Moreira et al., 2004) and the introduction of equivalent mutations in the yeast homologue Sen1 provokes termination defects both *in vivo* and *in vitro* (Leonaitė et al., 2017).

III.IV RNases H

Once formed, DNA:RNA hybrids are very stable, nevertheless they must be removed. Ribonuclease H (RNase H) plays a prominent role in this sense by being able to recognize and digest the RNA moiety of an heteroduplex (Hausen and Stein, 1970; Stein and Hausen, 1969). Two main types of RNase H exist, and at least one of them is found in most organisms. Eukaryotic RNase H1 enzymes function as single peptides (348 aa in yeast, 286 aa in human) and have a typical organisation consisting of highly conserved C- and N-terminal domains separated by a connection domain of variable length (Figure 19). The N-terminal of *S. cerevisiae* was the first one to be characterised and showed to contain two related motifs (Cerritelli and Crouch, 1995) that were later on named Hybrid Binding Domains (HBD) because of their ability to bind DNA:RNA hybrids with a 25-fold preference compared to the same RNA:RNA sequence (Nowotny et al., 2008). In most eukaryotes, the N-terminal of RNase H1 contains only one HDB and highly

RNase H1 is a modular enzyme containing a hybrid binding domain

Structural studies provided important clues on RNase H1 function

conserved aa, among which very few contact directly the substrate, while most of them are either structurally important or involved in nonspecific attraction of the duplex (Evans and Bycroft, 1999). The highly conserved FKKF motif interacts with the DNA strand and is placed in a shallow positively charged groove. W43 and F58 contact two deoxyribose rings of the DNA strand and having 2'-OH ribose groups in these positions would clash with the two aromatic aa, probably justifying the biased preference for hetero- over homoduplexes. The C-terminal domain (H domain) retains instead the catalytic activity, which requires at least four consecutive ribonucleotides (Figure 20). Structural studies showed that the hybrid sits in two shallow grooves in the enzyme with the 2'-OH of four consecutive ribose moieties interacting with one of the grooves while the DNA sits in the second one, with a phosphate fitting into a pocket for which a distortion of the DNA backbone is essential, a conformational change that RNA cannot accomplish (Nowotny et al., 2005). Thus, both strands of the heteroduplex contribute to the specificity of the enzyme. Several studies on *E. coli* RNase HI, clarified that four highly conserved carboxylic acid residues encompass the catalytic core and are essential for hydrolysis (Tadokoro and Kanaya, 2009). Because of its lack in sequence conservation and for being so variable in length, the connection domain is thought to provide flexibility allowing the N- and C-terminal regions to move rather freely in and around the substrates (Cerritelli and Crouch, 2009).

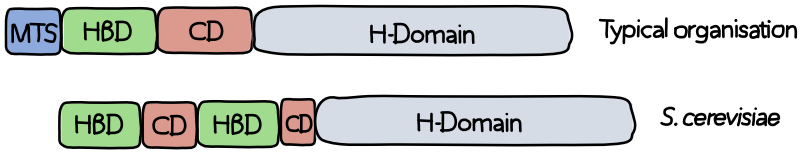


Figure 19. Domain organisation in RNase H1.
The 'typical' and the *S. cerevisiae* organisations are shown. Abbreviations: mitochondrial targeting sequence (MTS), hybrid binding domain (HBD), connection domain (CD) and RNase H domain (H-domain).

In metazoans the 5'-end of RNase H1 mRNA contains two alternative translation start sites. Synthesis from the first site embeds a Mitochondrial Targeting Sequence (MTS) into the protein that is responsible for its localisation in mitochondria where it is essential during development for the amplification of mitochondrial genome (Cerritelli et al., 2003).

Eukaryotic RNase H2 is composed of three subunits

Eukaryotic RNase H2 enzymes are heteromeric complexes, composed of three subunits named H2A, H2B and H2C (encoded in budding yeast by the *RNH201*, *RNH202* and *RNH203* genes respectively). *E. coli* RNase HII is instead a monomer, suggesting that eukaryotic RNases H2 evolved rather extensively and rapidly into a heterotrimer with the *S. cerevisiae* H2B and H2C subunits being the most distantly related to their human counterparts. Such disparity in aa sequence could explain why *in vivo* mutual complementation between the human and yeast factors cannot be obtained.

H2A is the catalytic subunit of RNase H2

Among the three subunits, H2A brings the catalytic function, while H2B and H2C act as scaffold proteins for H2A (Chon et al., 2009). Structural analysis of the enzyme from various organisms have revealed a great similarity between the active sites of *E. coli* RNase HII and the eukaryotic RNase H2, contained in the core domain of the H2A subunit (Figiel et al., 2011; Reijns et al., 2011; Rychlik et al., 2010; Shaban et al., 2010). RNase H2A active site has a conserved DEDD motif (Asp34, Glu35, Asp141 and Asp169 in human), coordinating metal ions, and a DSK (Asp67, Ser68 and Lys69 in human) motif, containing a loop placed in proximity to the active site and that is flexible in the absence of substrate. The hydroxyl group of the conserved Tyr of RNase H2A (Tyr210 in human and Tyr219 in yeast) interacts with the 2'-OH group of the rNMP to distort the substrate. Contrary to RNase H1, RNase H2 has the unique ability to bind and cleave single ribonucleoside monophosphates (rNMPs) embedded in the DNA (Eder et al., 1993).

In budding yeast, deletion of either one of the two RNase H enzymes does not alter normal growth and leads to a very slight sensitivity to DNA-damaging agents (Arudchandran et al., 2000). As already stated before, in more complex eukaryotes RNase H1 plays an important role

RNases H are likely redundant with one another and involved in R-loops metabolism

in mtDNA amplification and *Rnaseh1* null embryos undergo a development arrest. For what concerns its nuclear function instead, the specific role of RNase H1 remains overall debated. However, overexpression of nuclear RNase H1 has been extensively used to experimentally remove R-loops and it is so far perhaps the sole efficient tool to lower the cellular levels of R-loops (Skourti-Stathaki and Proudfoot, 2014). Moreover, deletion of both RNase H1 and H2 results in a strong increase in R-loops levels, indicating that the enzymes function at least to some extent redundantly in the removal of R-loops (Huertas and Aguilera, 2003; Wahba et al., 2011). Interestingly, RNase H1 is likely to be recruited via interaction with RPA, which also stimulates RNase H1 binding to DNA:RNA hybrids *in vitro* (Nguyen et al., 2017). In a study from 2016, Zimmer and Koshland took advantage of a set up specifically designed to test the frequency of loss of heterozygosity (LOH) as a readout of chromosome instability and showed that *rnh1Δ* cells behave as the wild-type counterpart (Zimmer and Koshland, 2016). Moreover, using ChIP, the authors demonstrated that Rnh1 can associate with R-loops across the genome, although it is only active at a small subset of those *loci*, and particularly at strong R-loop-forming ones (Zimmer and Koshland, 2016). Hence, it appears that Rnh1 may be negatively regulated in a constitutive manner and only turns into an active enzyme in response to R-loops stabilisation.

RNase H2 might play a role in rNMP excision

Mutations in subunits of RNase H2 have been associated with the Aicardi-Goutières syndrome (AGS), a neuro-inflammatory disease (Crow et al., 2006), chronic lymphocytic leukemia (CLL) and castration-resistant prostate cancer (CRPC) (Zimmermann et al., 2018). In virtue of its ability to recognise single rNMPs, RNase H2 was suggested to partake in the ribonucleotide excision repair (RER) pathway, by which rNMPs mistakenly incorporated during DNA synthesis are excised from the otherwise duplex DNA, which is followed by nicks resealing (Williams et al., 2016). Consistent with the idea that RNase H1 plays a minor role in physiological conditions, RNase H2 removal results in stronger genetic instability phenotypes in yeast (Conover et al., 2015; O'Connell et al., 2015; Zimmer and Koshland, 2016). Nevertheless, it remains still not very clear if the genetic instability derives from a

defective RER or R-loops accumulation. The generation of a separation-of-function mutant allele of *RNH202* that is specifically defective for the RER, hence named *RNH201*-RED, has suggested that the majority of genomic instability can be related to R-loop stabilisation (Chon et al., 2013), even though the expression of this allele does not rescue all of the genomic instability observed in cells lacking RNase H2 activity (Zimmer and Koshland, 2016). This idea that RNase H2 plays a major role in R-loops removal and that RNase H1 also contributes to R-loop removal but only in non-physiological conditions (i.e., when R-loops are stabilised) was also supported by analysing the association of Rnh1 to the chromatin in R-loops accumulating cells (Lockhart et al., 2019).

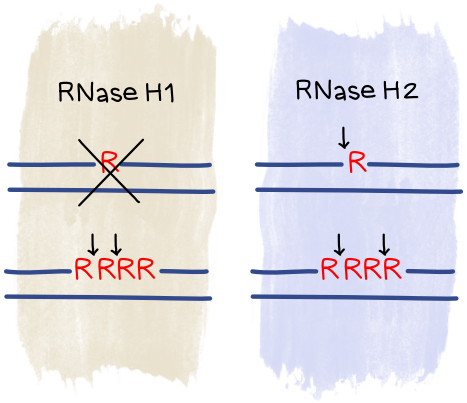


Figure 20. Cleaving patterns of RNases H.
A single ribonucleotide (red R) in a duplex DNA is cleaved by RNase H2 but not by RNase H1. Four consecutive ribonucleotide residues are cleaved differently by the two RNases H. Cleaving sites are indicated by black arrows.

III.V Mapping of R-loops

Understanding the formation dynamics, the function and the impact of R-loops in the different cellular processes requires sensitive and resolutive R-loop mapping methodologies. The study of R-loops has so far relied on three kinds of strategy (Figure 21): (i) the use of the

monoclonal S9.6 antibody that has a strong affinity for DNA:RNA hybrids (Boguslawski et al., 1986); (ii) the detection of the binding sites of a catalytically dead RNase H1 enzyme (Wu et al., 2001); (iii) the treatment with bisulfite (Yu et al., 2003).

Methods to measure overall R-loops levels

The overall assessment of R-loops levels in a given cell or condition is generally relying on the use of the S9.6 antibody. DNA:RNA hybrids can be detected on spotted nucleic acid preparations with the S9.6 antibody (S9.6 dot-blot). Appropriate controls are required, notably treatment with RNase H to ensure the disappearance of the S9.6 signal upon digestion and use of an **α**-DNA antibody in parallel with as a loading control. Global R-loop levels can also be detected by S9.6 immunofluorescence (IF), although increasing evidences revealed that a considerable amount of the signal detected is actually due to the residual affinity of the S9.6 antibody for dsRNA (Hartono et al., 2018; Phillips et al., 2013), therefore challenging quantification of the DNA:RNA hybrids. For instance, while R-loops are expected to be confined in the nucleus and in the mitochondria, the majority of the signal detected by IF is cytoplasmatic, and accordingly derived from RNA but not from DNA:RNA hybrids. Recommendations for appropriate controls have to be followed for these experiments (Chédin et al., 2021).

The two methods above-mentioned allow the assessment of the overall levels of R-loops, and are very useful for set-ups requiring a large amount of conditions, such as screenings. On the other side, no information is retrieved concerning the levels of R-loops at specific *loci*.

DRIP and R-ChIP

This is, instead, achieved through other techniques that generally rely on Immunoprecipitation (IP) or, less frequently, on the treatment with bisulfite.

IP is employed either using the S9.6 antibody (DNA:RNA Immunoprecipitation, DRIP) or a catalytically dead RNase H1 (R-ChIP), which is unable to digest the RNA moiety of hybrids, yet still able to bind efficiently heteroduplexes (Wu et al., 2001). Following IP, the recovered material can be used to perform quantitative PCR on specific *loci* of interest, or it can be sequenced to retrieve information at

the genome-wide level. Also in this case, the experiment is often coupled with *in vitro* RNase H treatments as negative controls. When immunoprecipitating a catalytically dead RNase H, the enzyme is often transiently overexpressed, but it has been proposed that the expression of this defective enzyme does not alter R-loop metabolism. Both DRIP and R-ChIP suffer from the typical limitations of ChIP, such as poor resolution and a general lack of strand-specificity, which can only be gained with more laborious steps (e.g., by only sequencing the RNA portion of the immunoprecipitated, DRIPc). However, and strikingly, the R-loops landscape generated via the two techniques is quite different. DRIP approaches produce signal principally along transcribed gene bodies, GC-skewed CpG island promoters and terminal genic regions, particularly for closely spaced genes (Sanz et al., 2016). On the contrary, R-ChIP signal has a clear bias in favour of G-rich *loci* linked to promoter-proximal pausing of RNAPII and is generally depleted from termination regions and poorly found in gene bodies, and even when so, only a subset of genes overlaps with the ones identified by the S9.6 antibody (Chen et al., 2017). These inconsistencies have been proposed to be due to technical variability between the two methods or to the fact that RNase H1-based strategies might only identify a specific subset of DNA:RNA hybrids (Chédin et al., 2021). For instance, it is possible that *in vivo* RNase H1 requires accessory factors to recognise R-loops, likely RPA. This would imply that for unclear reasons many R-loops escape RPA binding and therefore, RNase H1. An alternative explanation is that R-loops recognised by R-ChIP belongs to a specific class of short, RNAPII pausing-associated hybrids, which fails to be detected by the S9.6 antibody likely because lost during the DNA:RNA extraction (Chédin et al., 2021). It is worth to mention that a relevant difference between the two techniques is that DRIP begins with the nucleic acids extraction, which is afterwards incubated with the S9.6 antibody, while during R-ChIP the dead RNase H1 is cross-linked *in vivo* and then immunoprecipitated. Consequently, during DRIP some R-loops might be too unstable to survive the extraction, or even some might form *ex vivo*, even though the latter is

Genome-wide maps generated via DRIP and R-ChIP differ from each other

probably a minor concern as the process is unfavoured by strong energy barriers.

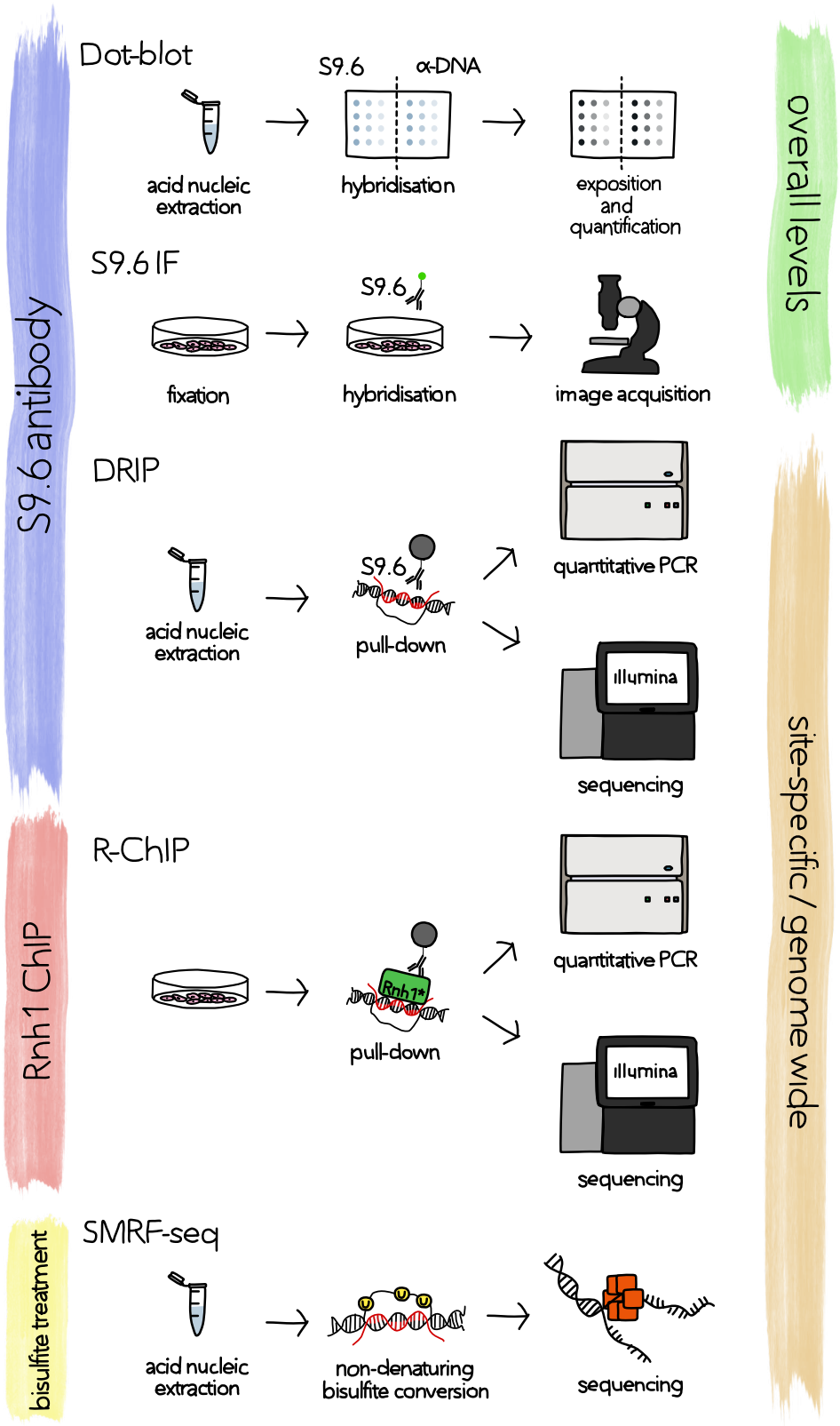


Figure 21. Current methods for R-loops detection.
Overview of the main steps for each of the technique currently used for R-loops detection.

SMRF-seq

The last technique available for the mapping of R-loops is the Single Molecule R-loop Footprinting (SMRF). The approach takes advantage of the presence of ssDNA in R-loops for bisulfite-induced deamination of exposed cytosines. *Bona fide* R-loops are considered for those regions where C-to-T conversion is strand-specific and sensitive to RNase H treatment (Chédin et al., 2021; Yu et al., 2003). The main advantage of the method is to allow a precise view of individual R-loop footprints, thus allowing to show that in the same *locus* individual R-loops can differ from one another for their 3′ and 5′ borders. On the other hand, the method remains less suited for genome-wide studies compared to the DRIP and R-ChIP.

Interestingly, a recent study focusing on 24 *loci* prone to form R-loops showed a strong agreement between S9.6 and SMRF signals (Malig et al., 2020).

AIM OF THE WORK

The evolutionary conserved helicase Sen1 plays a prominent role in limiting and resolving TRCs. On the one hand, as component of the Nrd1-Nab3-Sen1 (NNS) complex, Sen1 is an essential transcription termination factor, responsible for dislodging the transcription apparatus from thousands of non-coding genes (i.e., Cryptic Unstable Transcripts or CUTs). Sen1 has been proposed to prevent R-loops accumulation by virtue of its DNA:RNA unwinding activity. Currently, the specific mechanisms by which Sen1 affects R-loops levels and/or TRCs is unknown, and previous works have been performed with Sen1 mutants that strongly also affect its transcription termination function.

During my PhD I focused on the dissection of the role of Sen1 in genome stability, aiming at separating the functions of Sen1 in non-coding RNA genes termination and at R-loops and/or TRCs.

RESULTS

I. Sen1 is recruited to replication forks via Ctf4 and Mrc1 and promotes genome stability.

The following manuscript describes the binding of Sen1 to the replisome and the preliminary characterisation of the *sen1-3* allele.

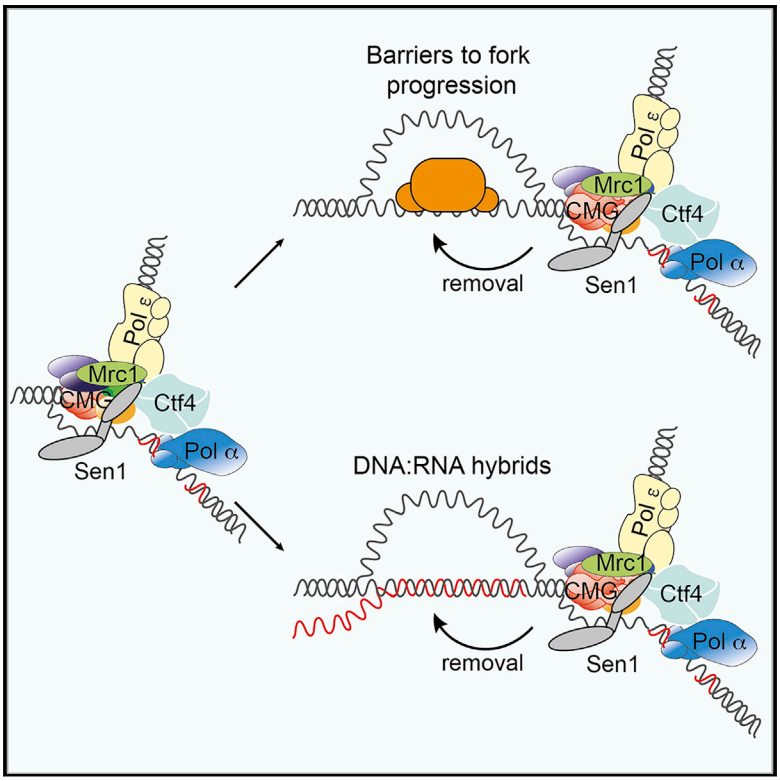
My contribution to this work was to investigate the presence of transcription termination defects in *sen1-3* cells, which I performed by RT-qPCR and RNAPII CRAC.

Report

Cell Reports

Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability

Graphical Abstract



Authors

Rowin Appanah, Emma Claire Lones, Umberto Aiello, Domenico Libri, Giacomo De Piccoli

Correspondence

g.de-piccoli@warwick.ac.uk

In Brief

Appanah et al. identify the transcription termination helicase Sen1 as a bona fide component of the replisome. Sen1 binds the replisome via its N-terminal domain and Ctf4 and Mrc1. The allele *sen1-3* breaks this interaction without affecting transcription termination. *sen1-3* cells show sensitivity to R-loops levels and increased genomic instability.

Highlights

- The N-terminal domain of Sen1 mediates replisome association
- Ctf4 and Mrc1 bind to Sen1 and promote its recruitment to the replisome
- The *sen1-3* allele abrogates replisome binding, but not transcription termination
- *sen1-3* cells are sensitive to high levels of R-loops and defects of S phase checkpoint

Appanah et al., 2020, Cell Reports *30*, 2094–2105
February 18, 2020
https://doi.org/10.1016/j.celrep.2020.01.087

CellPress

OPEN ACCESS
CellPress

Cell Reports
Report

Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability

Rowin Appanah,[1] Emma Claire Lones,[1] Umberto Aiello,[2] Domenico Libri,[2] and Giacomo De Piccoli[1,3,*]
[1]Warwick Medical School, University of Warwick, CV4 7AL Coventry, UK
[2]Institut Jacques Monod, CNRS, UMR7592, Université Paris Diderot, Paris Sorbonne Cité, Paris, France
[3]Lead Contact
*Correspondence: g.de-piccoli@warwick.ac.uk
https://doi.org/10.1016/j.celrep.2020.01.087

SUMMARY

DNA replication and RNA transcription compete for the same substrate during S phase. Cells have evolved several mechanisms to minimize such conflicts. Here, we identify the mechanism by which the transcription termination helicase Sen1 associates with replisomes. We show that the N terminus of Sen1 is both sufficient and necessary for replisome association and that it binds to the replisome via the components Ctf4 and Mrc1. We generated a separation of function mutant, *sen1-3*, which abolishes replisome binding without affecting transcription termination. We observe that the *sen1-3* mutants show increased genome instability and recombination levels. Moreover, *sen1-3* is synthetically defective with mutations in genes involved in RNA metabolism and the S phase checkpoint. *RNH1* overexpression suppresses defects in the former, but not the latter. These findings illustrate how Sen1 plays a key function at replication forks during DNA replication to promote fork progression and chromosome stability.

INTRODUCTION

The maintenance of genome stability requires the complete and faithful duplication of DNA in every cell cycle. Yet several obstacles impede the progression of replication forks (RFs), and these must be removed to avoid stalling and increased chromosome instability. A significant barrier to RF progression is transcription. First identified in bacteria, collisions between RFs and transcription bubbles also represent a major obstacle for DNA synthesis in eukaryotes, leading to defects in chromosome maintenance and an increase in levels of recombination (Liu and Alberts, 1995; Helmrich et al., 2011, 2013; Prado and Aguilera, 2005; Kim et al., 2010; Hamperl et al., 2017; Tran et al., 2017). In order to complete the full duplication of the chromosomes, replisomes must therefore overcome transcriptional barriers, removing both the DNA-bound RNA polymerase subunits and any DNA:RNA hybrids formed during transcription. These hybrids, usually limited to eight base pairs, occur naturally during RNA transcription and are typically removed when the RNA polymerase is disengaged from the DNA (Aguilera and García-Muse, 2012; Westover et al., 2004).

At specific chromosomal loci, extended DNA:RNA hybrids can also form behind the site of RNA synthesis, through the re-annealing of nascent RNA to the template DNA and the displacement of the non-template DNA. These structures, named R-loops, form preferentially at highly transcribed genes with a high GC skew and can extend up to 1 kb in higher eukaryotes (Aguilera and García-Muse, 2012; Skourti-Stathaki et al., 2014). Formation of R-loops is favored by head-on collisions between RFs and actively transcribing complexes (Hamperl et al., 2017; Lang et al., 2017), and their non-physiological accumulation, coupled to chromatin modification, is deleterious for genome stability (García-Pichardo et al., 2017). Several pathways minimize the formation and stability of R-loops. For instance, the promotion of transcription processivity (Hazelbaker et al., 2013), transcription termination (Kim et al., 2004; Luke et al., 2008), timely processing, export or degradation of nascent mRNA (Huertas and Aguilera, 2003; Pfeiffer et al., 2013), or preventing torsional stress that arises during transcription (El Hage et al., 2010, 2014) all minimize R-loops' levels. Nevertheless, once formed, R-loops must be removed. A key role in R-loop removal is fulfilled by the RNase H enzymes that specifically digest RNA molecules within DNA:RNA hybrids (Cerritelli and Crouch, 2009). In addition, several helicases can unwind DNA:RNA hybrids *in vitro*, including Sgs1 (Chang et al., 2017) and Pif1 (Boulé and Zakian, 2007). One such helicase, Sen1, is believed to play an essential role in the removal of R-loops from the DNA in yeast (Mischo et al., 2011).

Sen1 is an Upf1-like helicase that plays a key role in transcription termination (Jankowsky, 2011; Steinmetz et al., 2006; Ursic et al., 1997; Porrua and Libri, 2013). Sen1 binds to the free 5′ ends of either RNA or DNA substrates and unwind both double-stranded DNA (dsDNA) and DNA:RNA hybrids (Han et al., 2017; Leonaitė et al., 2017; Martin-Tumasz and Brow, 2015; Porrua and Libri, 2013). *In vitro* analysis shows that Sen1 has high activity but limited processivity on DNA:RNA hybrid substrates (Han et al., 2017). Mechanistically, when Sen1 engages with nascent RNA exiting from a stalled RNA polymerase II (RNAPII), the helicase seemingly exerts a force on the polymerase to "push" it, either overcoming the stalling of RNAPII or disengaging it from the template DNA (Porrua and Libri, 2013; Han et al., 2017). *In vivo* data also suggest that Sen1 is capable of removing RNAPII from the DNA it is bound to, thus terminating transcription (Steinmetz et al., 2006; Schaughency et al., 2014; Hazelbaker et al., 2013). In fact, a mutation in the catalytic domain of Sen1 (*sen1-1*) confers defects in transcription termination at non-permissive temperatures, leading to extensive readthrough

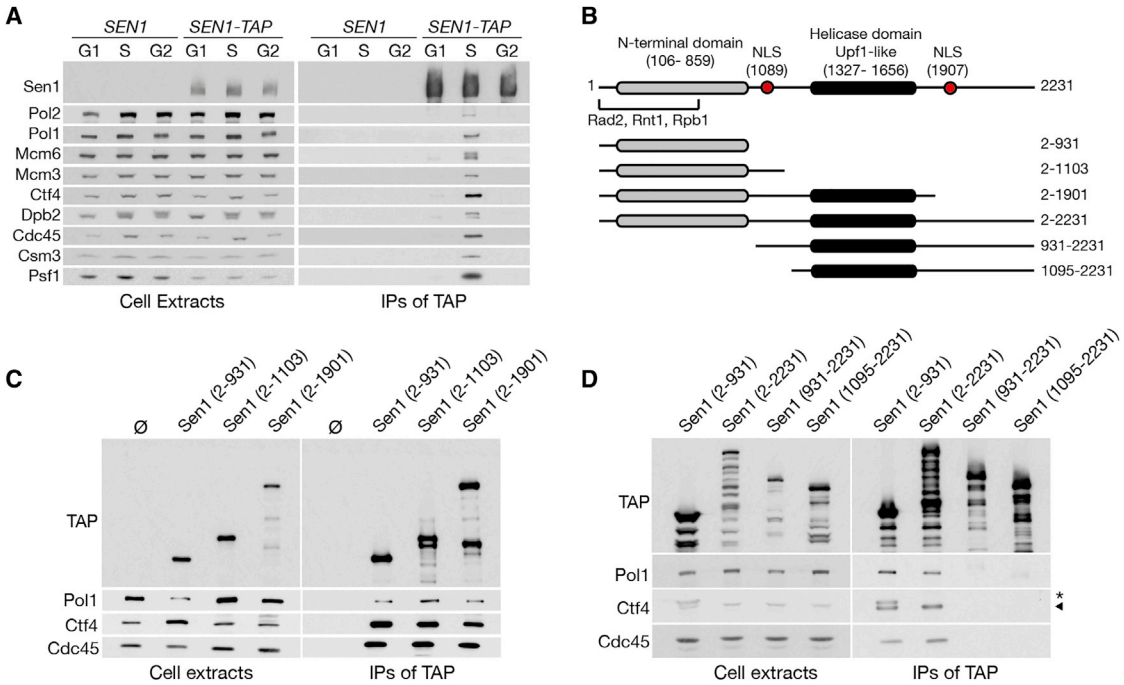


Figure 1. Sen1 Interacts with the Replisome during S Phase through Its N-Terminal Domain

(A) *SEN1* or *SEN1-TAP* cells were arrested in G1, harvested immediately, or released for either 30 min (S phase) or 60 min (G2 phase). Cell extracts and IP material were analyzed by immunoblotting (IB).

(B) Schematic of Sen1 constructs used.

(C) TAP-tagged fragments of Sen1, IPed from cells in S phase, were analyzed by IB.

(D) TAP-tagged fragments of Sen1 were analysed as above, except 4× cells were used for the IP of the fragments containing the last 330 C-terminal amino acids.

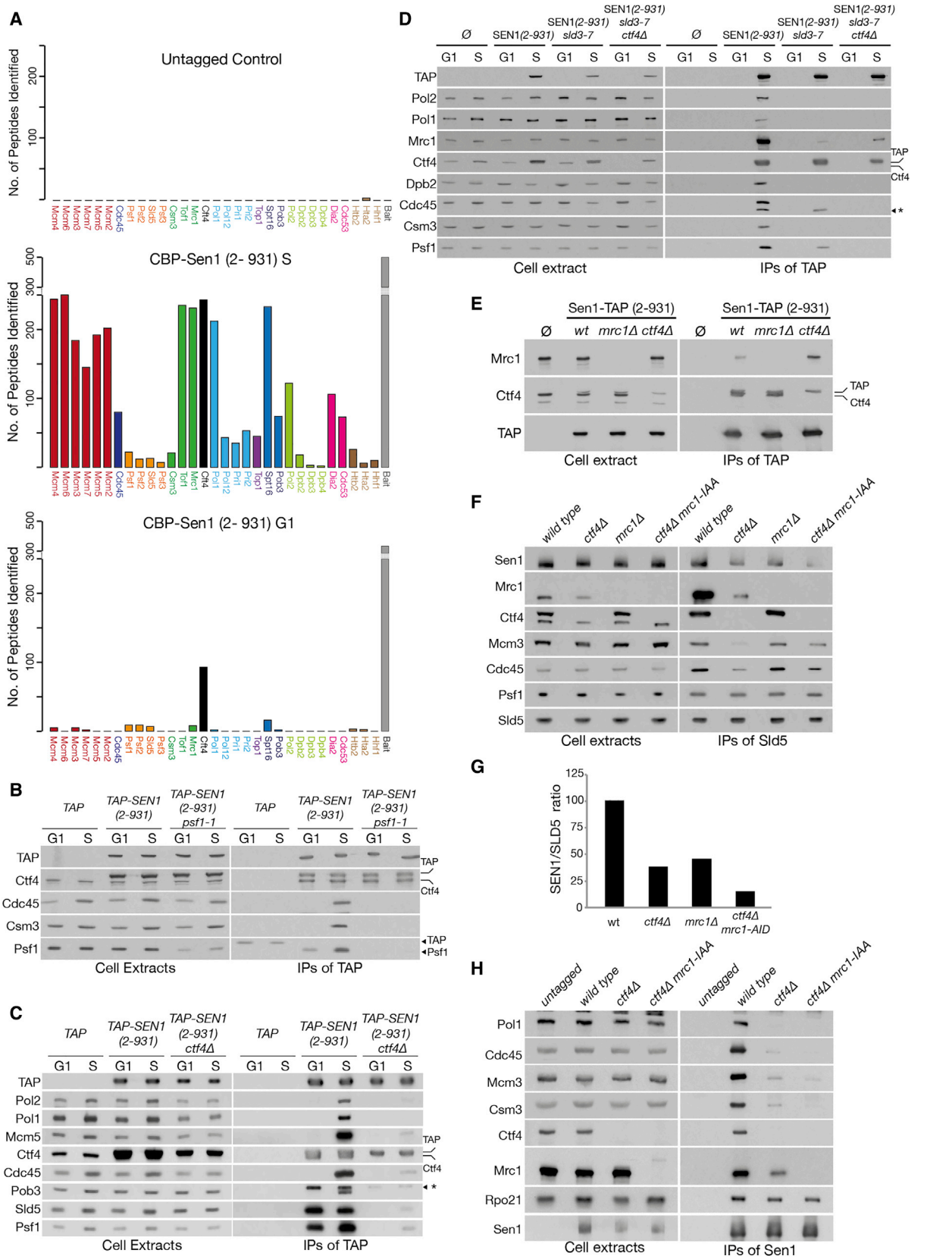
of several transcription units (Steinmetz et al., 2006), accumulation of R-loops, and increased recombination (Mischo et al., 2011). Because of these defects, the viability of *sen1-1* cells depends on several repair factors (Mischo et al., 2011; Alzu et al., 2012). Moreover, depletion of Sen1 leads to slow DNA replication and the accumulation of abnormal structures on 2D gels (Alzu et al., 2012; Brambati et al., 2018).

Given its relatively low abundance and processivity (Mischo et al., 2018; Han et al., 2017), Sen1 needs to be recruited at, or close to, sites where it can enact its biological function. Sen1 is recruited to the termination sites of cryptic-unstable transcripts (CUTs) and small nucleolar RNAs (snoRNAs) by binding to Nab3 and Nrd1, which both dock onto nascent RNA (Arigo et al., 2006; Porrua et al., 2012; Creamer et al., 2011). Nrd1 also interacts with $Rpo21^{Rpb1}$ (the largest subunit of RNAPII) early in the transcription cycle (Vasiljeva et al., 2008), thus restricting Sen1-dependent termination to short transcription units (Gudipati et al., 2008). Sen1 also promotes termination of some genes downstream of the polyadenylation site, acting with Rat1 (Mischo et al., 2011; Rondón et al., 2009), possibly by directly binding Rpo21 via its N-terminal domain (Chinchilla et al., 2012). Finally, it is likely that Sen1 is recruited at other genomic sites in a transcription-independent fashion. The human ortholog of Sen1 (Senataxin) co-localizes with 53BP1 to sites of DNA damage in a checkpoint-dependent manner (Yüce and West, 2013). Moreover, in *S. cerevisiae*, Sen1 co-localizes with replisome components and sites of bromodeoxyuridine (BrdU) incorporation (Alzu et al., 2012). However, the mechanism through which Sen1 is recruited at RFs has yet to be described. The significance of recruiting Sen1 to RFs is also poorly understood, as it has been impossible thus far to determine whether the defects in DNA replication upon inactivation of Sen1 are an indirect consequence of deregulated transcription termination, of a failure in R-loop removal, or the direct result of an important function of Sen1 at RFs. Here, we show that Sen1 binds the replisome during S phase through its N-terminal domain, map its binding site, generate a mutant that breaks this interaction, and explore the consequences of the loss of the helicase from RFs on chromosome stability.

RESULTS

Sen1 Interacts with the Replisome via Its N-Terminal Domain

The replisome is a complex and dynamic machine that relies on multiple interactions between its constituent proteins (Bell and Labib, 2016; Burgers and Kunkel, 2017). As part of a mass spectrometry (MS) screen to identify factors transiently or weakly associated to the core replisome, we observed that Sen1 co-purifies with the CMG helicase in *S. cerevisiae* (Figure S1A). To verify the MS data, we immunoprecipitated (IPed) Sen1 from extracts of yeast cells synchronized in G1, S, and G2. We observed that Sen1 interacted with replisome components only in S phase (Figure 1A). Immunoprecipitation (IP) of the GINS component



(legend on next page)

Sld5 corroborated this observation (Figure S1B). Sen1 interacts with replisomes independently of either Nrd1 or Nab3 (Figures S1C and S1D) and independently of ongoing transcription (Figures S1E and S1F), as previously observed (Alzu et al., 2012). To further explore this interaction and its biological function, we mapped the interaction sites both in the replisome and Sen1.

Sen1 contains an extended N-terminal domain and an essential and conserved helicase domain (Leonaitė et al., 2017). To identify a region of Sen1 that is sufficient for binding replisomes, we generated TAP-tagged constructs of Sen1, expressed under an inducible *GAL1* promoter (Figure 1B). All fragments containing the helicase domain folded correctly and rescued *sen1-1* lethality at non-permissive temperatures, despite constructs containing the last 330 amino acids of the protein being highly labile (Figures S1G and S1H). We then assessed the ability of the various fragments to interact with the replisome and observed that the N-terminal domain (residues 2–931) of Sen1 was both sufficient and necessary for association with replisomes (Figures 1C and 1D). Similarly, Sen1 (2–931) co-precipitated specifically with replisomes isolated from S phase cells by IP of Mcm3 (a subunit of the CMG helicase) (Figures S1I and S1J). Thus, Sen1 (2–931) contains an interaction site for replisome components.

Sen1 Binding to the Replisome Depends on Ctf4 and Mrc1

To identify specific proteins to which Sen1 binds within the replisome, we compared the G1 and S phase interactome of Sen1 (2–931) via MS analysis. As expected, Sen1 (2–931) IPed with replisomes in S phase (Figure 2A). Interestingly, Ctf4 and GINS co-purified with the bait in G1 as well. This was confirmed by immunoblotting (Figures 2B and 2C). Because Ctf4 and GINS interacts throughout the cell cycle (Gambus et al., 2009), we next analyzed whether Sen1 binds preferentially to one of the components. The interaction between Ctf4 and Sen1 in G1 was unaffected by inactivating GINS via the *psf1-1* allele (Figure 2B; Takayama et al., 2003), but GINS no longer IPed with Sen1 (2–931) in G1 in the absence of Ctf4 (Figure 2C). These data indicate that Sen1 (2–931) binds to Ctf4 in the absence of other replisome components.

Interestingly, Sen1 (2–931) retained some affinity to the replisome in the absence of Ctf4 (Figure 2C, right panel), independently of DNA (Figure S2A). This suggests that Sen1 interacts with at least another subunit of the replisome. To screen for such factors, we analyzed whether any component of the replisome binds to Sen1 (2–931) in cells progressing into S phase in the absence of origin firing. We used *td-sld3-7* cells that cannot initiate chromosome replication at 37°C following inactivation and degradation of td-Sld3-7 (Kamimura et al., 2001; Kanemaki and Labib, 2006; Figure S2B). In control cells, Sen1 (2–931) co-purified with all tested replisome components in S phase (Figure 2D). In *td-sld3-7* cells, Sen1 IPed predominantly with Ctf4 and GINS but also weakly with the replisome component Mrc1. Strikingly, Sen1 (2–931)'s affinity for Mrc1 increased in a td-*sld3-7 ctf4Δ* background. We confirmed this in cells arrested in G1 as well (Figure 2E). These observations suggest that both Ctf4 and Mrc1 are binding partners of Sen1 in the replisome. Deletion of either replisome component leads to a decrease in replisome association to Sen1, even following crosslinking to capture weak interactions (Figures S2C and S2D). Because *ctf4Δ mrc1Δ* cells are inviable (Gambus et al., 2009), we generated a *ctf4Δ mrc1-AID* strain, with the auxin-degron fused to Mrc1 (Nishimura et al., 2009) to allow rapid depletion of the protein. The association of Sen1 with the replisome was greatly reduced, although not entirely abolished, in cells with no Ctf4 and Mrc1 (Figures 2F–2H). These data indicate that, although other accessory binding partners might exist within the replisome, Sen1 mainly binds via Ctf4 and Mrc1.

sen1-3 Fails to Bind the Replisome and Is Sensitive to Increased Levels of DNA:RNA Hybrids

Deletion of the N-terminal domain of Sen1 causes pronounced defects in cell growth (Figure S3A). Thus, to investigate the role of Sen1 at RFs, we sought to generate a separation of function allele that is specifically defective for binding to replisomes. By generating truncations of the N-terminal domain, we identified that Sen1 (410–931) was the fragment with the highest affinity for replisomes although Sen1 (622–931) was the smallest construct still able to bind (Figures 3A–3C). By comparison with yeast orthologs of Sen1 (Figure S3B), we identified conserved residues within this region and targeted them for mutagenesis, creating hemagglutinin (HA)-tagged alleles of *SEN1* that were expressed under the strong *ACT1* promoter in *sen1Δ* cells. All the tested mutations supported cell growth, but one allele, combining mutations W773A E774A W777A (henceforth referred to as *sen1-3*) was uniquely defective for interaction with replisomes (Figures 3D and S3C). Similar results were obtained when the *sen1-3* mutation was

Figure 2. Sen1 Binds the Replisome Components Ctf4 and Mrc1

(A) MS analysis of the proteins co-purifying with Sen1 (2–931) was conducted in S and G1 phases.

(B) IB analysis of the proteins IPed with Sen1 (2–931) and an empty control in strains carrying the *PSF1* or *psf1-1* allele. Cells were arrested in G1, shifted to 37°C for 1 h (G1), and then released into S phase for 20 min at 37°C (S).

(C) Sen1 (2–931) binding of GINS in G1 depends on Ctf4. IB analysis of the proteins IPed with Sen1 (2–931) and an empty control, with or without *CTF4*. Cells were arrested in G1 and released in S phase for 20 min at 30°C. Ctf4 and TAP-Sen1 (2–931) have similar sizes and run closely in gel electrophoresis.

(D) IB analysis of the proteins interacting with TAP-Sen1 (2–931) in the presence or absence of origin firing and *CTF4*. Cells were treated as described in Figure S2B. G1 samples were collected before galactose induction.

(E) Wild-type, *mrc1Δ*, or *ctf4Δ* cells expressing TAP-Sen1 (2–931) were arrested in G1. IB analysis of cell extracts and IPs is shown.

(F) Wild-type, *ctf4Δ*, *mrc1Δ*, and *ctf4Δ mrc1-AID* strains were arrested in G1, treated for 1 h with 0.5 mM auxin indole-3-acetic acid (IAA) final concentration, and released in S phase. IB analysis of cell extracts and IPs is shown.

(G) Quantification of the relative signal of Sen1-9MYC versus the TAP-Sld5 signal, normalized against the wild type.

(H) Experiments were conducted as in (F). Wild-type, *ctf4Δ*, and *ctf4Δ mrc1-AID* strains, carrying an untagged or a *SEN1-TAP* allele, were used. Asterisk indicates a non-specific band.

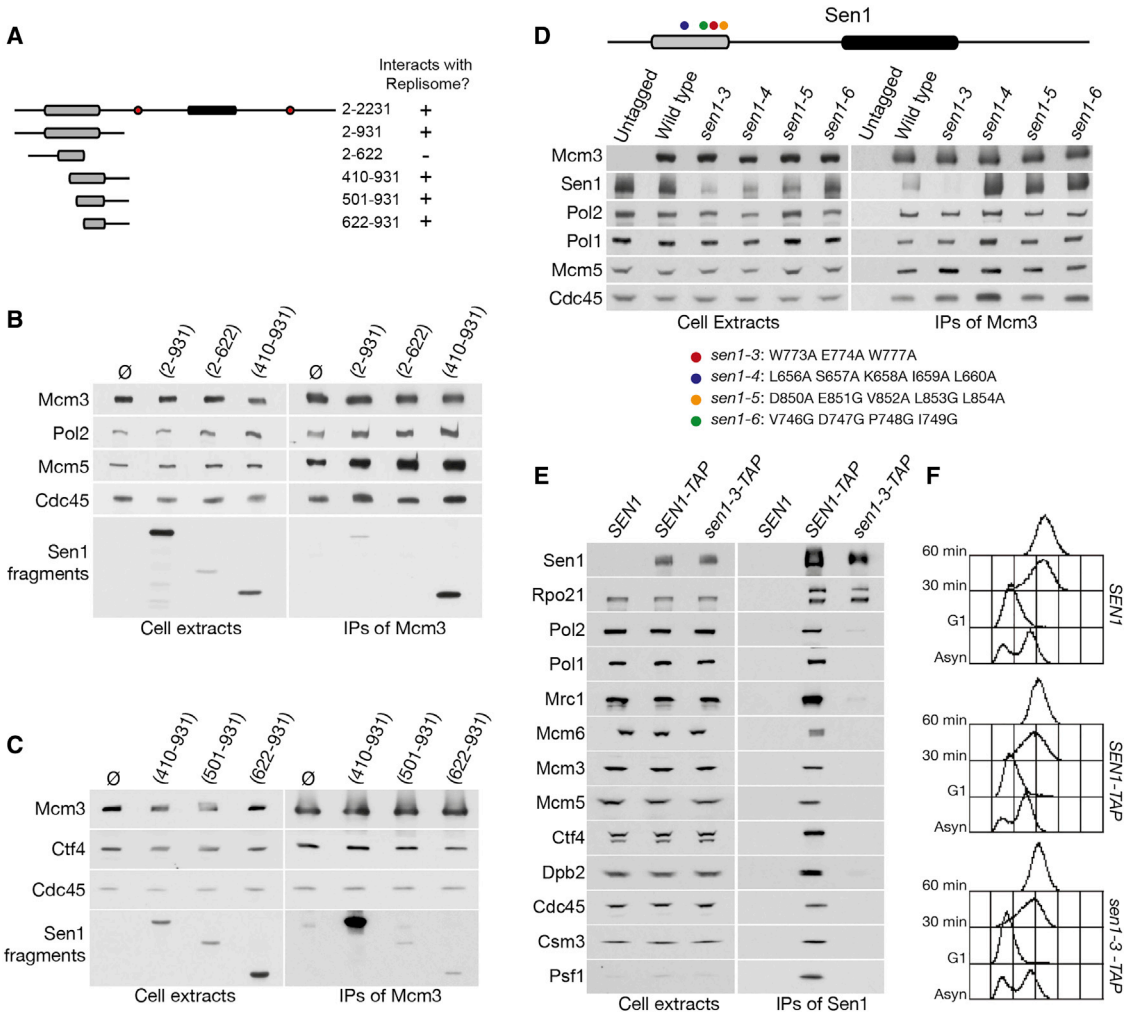


Figure 3. Sen1-3 Does Not Interact with the Replisome

(A) Summary of the ability of N-terminal fragments of Sen1 to interact with the replisome.

(B) Cells carrying different *GAL1-3HA-SEN1* fragments and a *TAP-MCM3* allele were arrested in G1 and released into S phase. The samples were then used for IPs.

(C) Sen1 fragments were analysed as in (B).

(D) Cells carrying *ACT1-3HA-SEN1* wild-type or mutated alleles at an ectopic locus were synchronously released into S phase. IB analysis of cell extracts and IPs is shown.

(E) Cells carrying a *SEN1*, *SEN1-TAP*, or *sen1-3-TAP* allele were arrested in G1 and released into S phase. IB analysis of cell extracts and IPs is shown.

(F) Fluorescence-activated cell sorting (FACS) samples for the experiment in (E).

introduced at the endogenous *SEN1* locus (Figures 3E and 3F), even following crosslinking (Figure S3D). Importantly, Sen1-3 retained wild-type affinity for RNAPII (Rpo21). Hence, *sen1-3* is an allele that abrogates the interaction between Sen1 and replisomes.

Next, we assessed whether the *sen1-3* mutation affects transcription termination, similarly to *sen1-1* cells (Mischo et al., 2011). We assayed the efficiency of termination at two model Nrd1-Nab3-Sen1 target genes, coding for a snoRNA (*SNR13*) and a CUT (*NEL025c*) (Thiebaut et al., 2006; Ursic et al., 1997). Because termination defects lead to longer RNAs that can be targeted by the nonsense-mediated decay, strains lacking *UPF1* were also tested (Culbertson and Leeds, 2003). Cells with the *sen1-3* allele presented no defects in transcription termination, unlike *sen1-1* at 37°C (Figures 4A and S3E). Defects in transcription termination were also analyzed genome-wide by mapping the distribution of RNAPII via the sequencing of nascent RNAs using CRAC (crosslinking and analysis of cDNAs) (Granneman et al., 2009; Candelli et al., 2018). Metagene analysis using a set of validated CUTs (Table S1) shows very similar RNAPII profiles between *SEN1* and *sen1-3* cells, although a clear general termination defect is observed upon depletion of Nrd1 (Figures 4B and S3F). These data indicate that *sen1-3* is proficient in terminating RNAPII transcription.

We then analyzed how the loss of Sen1 from replisomes affects cells. *SEN1* and *sen1-3* cells displayed comparable cell

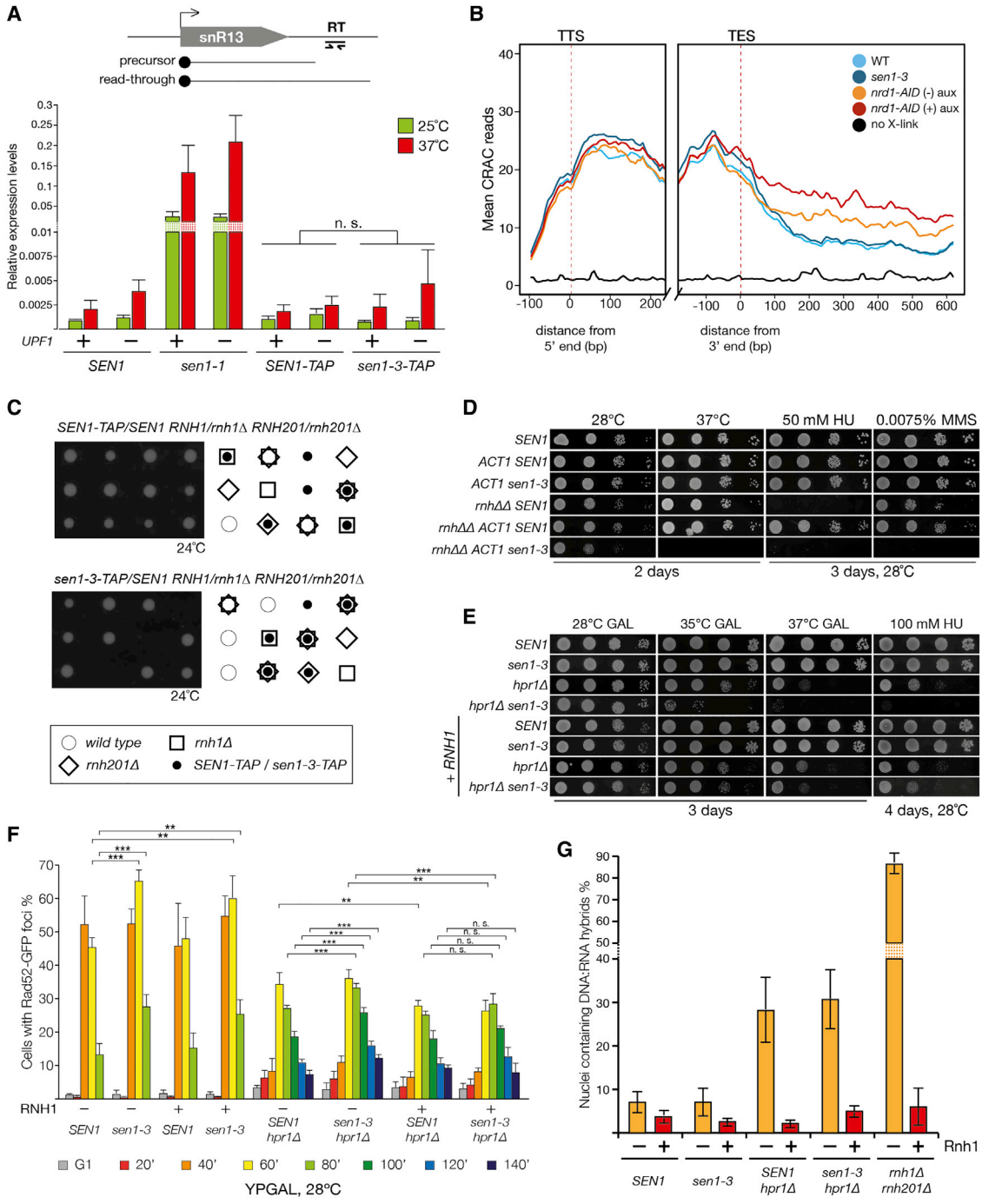


Figure 4. The *sen1-3* Allele Is Proficient in RNAPII Termination but Is Essential in the Absence of RNase H Activity

(A) *sen1-3* cells are proficient for transcription termination. qRT-PCR analysis of RNAs derived from the strains indicated is shown. The ratio of the readthrough fraction (position RT) over the total amount of *SNR13* RNA is shown (triplicate biological repeats). n.s., not significant.

(B) Metagene analysis of RNAPII density detected by CRAC on CUTs. Average read counts are plotted on regions aligned to both the transcription start site (TSS) (left) and the transcript end site (TES) (right) of the CUTs (reads count in Table S1). The profiles of RNAPII density following Nrd1 depletion (*nrd1-AID* + auxin) are included for comparison (dataset from Candelli et al., 2018). *nrd1-AID* strain behaves as a hypomorphic allele.

(C) Examples of the meiotic progeny of the indicated diploids strains are shown.

(D) Serial dilution spotting of the indicated strains is shown. *rnh1Δ rnh201Δ* is abbreviated as *rnhΔΔ*.

(E) Serial dilution spotting of the indicated strains is shown. Cells (+*RNH1*) carry *GAL-RNH1* inserted ectopically.

(legend continued on next page)

growth kinetics and sensitivity to both hydroxyurea (HU) and methyl methanesulfonate (MMS). One possibility might be that Sen1 at RFs is redundant with the enzymatic activity of other factors, such as the RNase H1 and H2 enzymes. We crossed *rnh1Δ rnh201Δ* cells with *SEN1 or sen1-3* strains and analyzed their meiotic progeny. Although single deletion of either *RNH1* or *RNH201* combined with *sen1-3* did not present any synthetic defects, *sen1-3 rnh1Δ rnh201Δ* cells were inviable (Figure 4C), similarly to *rnh1Δ rnh201Δ sen1-1* mutants (Figure S3G). Overexpression of *sen1-3* under the strong *ACT1* promoter suppresses the synthetic lethality of *sen1-3* with *rnh1Δ rnh201Δ*, suggesting that higher levels of Sen1 activity can compensate for lack of the specific replisome-tethering mechanism. Yet these cells display growth defects at 37°C, with cells accumulating in G2/M and triggering checkpoint activation (Figures S3H–S3J). Moreover, *ACT1-sen1-3* is unable to suppress the hyper-sensitivity of *rnh1Δ rnh201Δ* to HU and is synthetic defective for MMS sensitivity (Figure 4D). Altogether, these findings suggest that Sen1 at RF might either be redundant with RNases H1 and H2 or become essential to deal with the DNA:RNA hybrids accumulating in the absence of RNase H.

To explore whether increased levels of DNA:RNA hybrids lead to synthetic defects in *sen1-3* cells, we generated *hpr1Δ sen1-3* cells. Hpr1 is a component of the THO complex involved in the processing and export of mRNA (Chávez et al., 2000). *hpr1Δ* mutants accumulate R-loops and show defects in transcription elongation (García-Benítez et al., 2017; Chávez and Aguilera, 1997; Chávez et al., 2000). *hpr1Δ sen1-3* double mutants showed growth defects at higher temperatures and increased sensitivity to replication stress (Figure 4E). To explore whether defects arise during DNA replication, we analyzed the kinetics of Rad52 foci formation in cells released in S phase. The experiment was conducted at permissive temperatures (28°C) as *hpr1Δ* cells failed to synchronously bud at 35°C and 37°C. We observed that *sen1-3* causes a small but statistically significant increase in recombination in late S phase, although *hpr1Δ sen1-3* cells showed synthetic defects and an increase in recombination (Figures 4F and S4A–S4D). Interestingly, the increased rates of recombination and growth defects in *hpr1Δ sen1-3* cells were suppressed by overexpression of *RNH1* (Figures 4E and 4F), thus suggesting that DNA:RNA hybrids are toxic in these mutants.

To directly test the levels of DNA:RNA hybrids, we visualized them in chromosome spreads (Wahba et al., 2011). As previously observed, both *rnh1Δ rnh201Δ* and *hpr1Δ* mutants showed high levels of DNA:RNA hybrids (Figures 4G and S4E; Chan et al., 2014). Surprisingly, we did not observe any increase in the levels of DNA:RNA hybrids in *hpr1Δ sen1-3* cells. Similar results were observed by slot-blot analysis (Figure S4F). Given that phenotypic suppression by RNase H overexpression is accepted as a marker for R-loops, these results suggest that the suppression of *hpr1Δ sen1-3* by overexpression of *RNH1* might occur by removing short or labile DNA:RNA hybrids, not readily detectable by the S9.6 antibody used in our analysis.

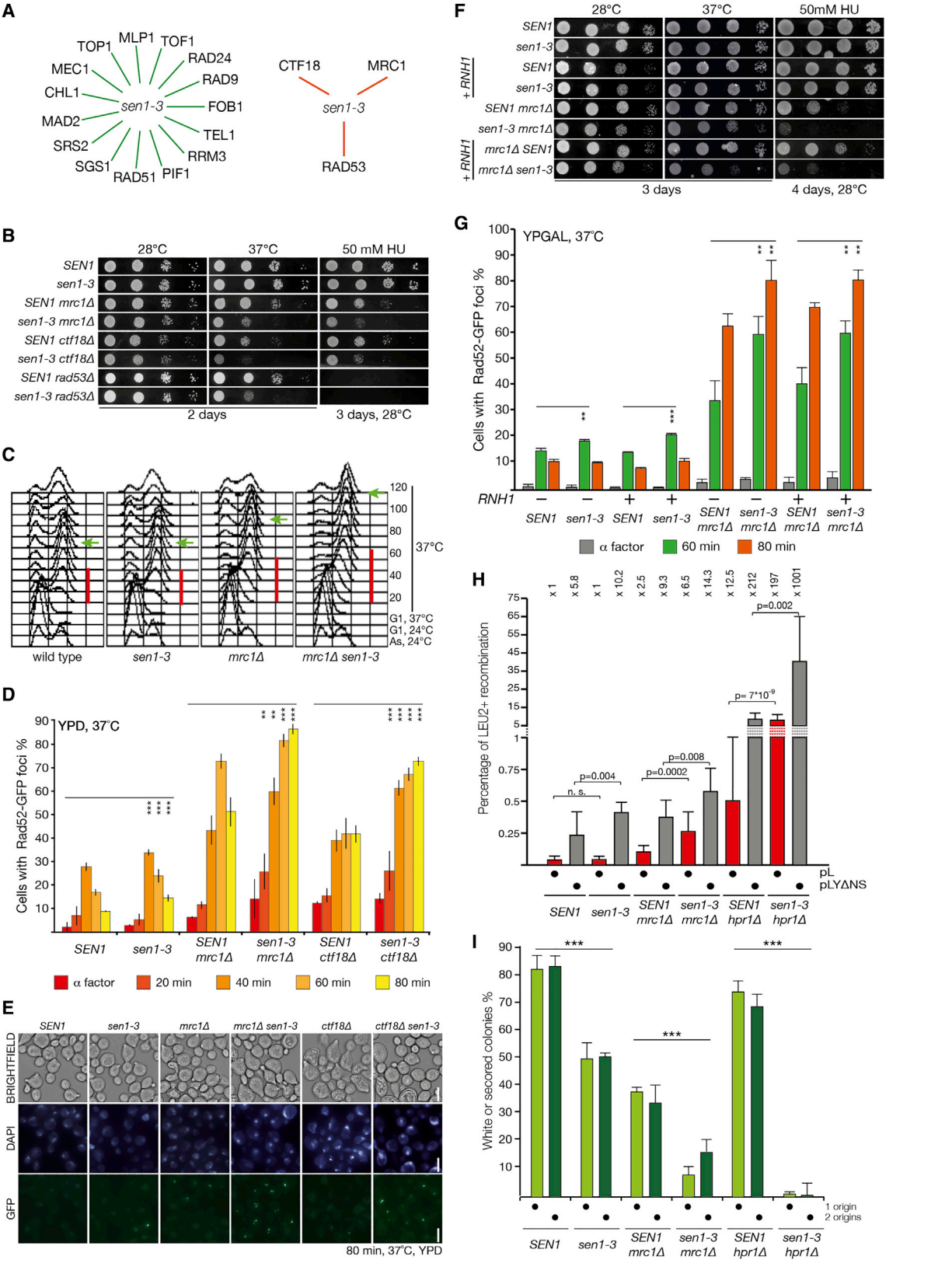
sen1-3 Cells Are Defective in Replication Fork Progression and Genome Stability in the Absence of *MRC1*

Because both *sen1-1* and *sen1-3* are synthetically lethal in the absence of *RNH1* and *RNH201*, we wanted to explore whether other pathways, essential for maintaining cell viability in *sen1-1* (Alzu et al., 2012; Mischo et al., 2011), are also important in *sen1-3*. Only a subset of deletion mutants described to negatively affect viability in *sen1-1* cells showed robust defects in cell viability in *sen1-3* cells (summarized in Figure 5A). Namely, we observed temperature sensitivity and increased sensitivity to DNA-damaging agents when *sen1-3* was crossed with either *mrc1Δ*, *ctf18Δ*, or *rad53Δ sml1Δ* (Figure 5B).

Mrc1, Ctf18, and Rad53 are key components of the S phase checkpoint, and all three mutants confer temperature sensitivity in *sen1-3* cells. To further explore the defects of *mrc1Δ sen1-3* mutants, we analyzed the DNA replication dynamics of cells arrested in G1 and then released in S phase at 37°C. The *mrc1Δ sen1-3* cells show a delay during DNA replication and accumulation of cells arrested in G2/M (Figure 5C). Correspondingly, we observed an increase in Rad52-GFP foci accumulating during the later stages of DNA replication, both in the *mrc1Δ sen1-3* and *ctf18Δ sen1-3* cells released in S phase at 37°C (Figures 5D and 5E). In addition, *mrc1Δ sen1-3* and *ctf18Δ sen1-3* cells showed an increase in cells carrying multiple foci of Rad52 (Figure S5A). Similar to what is seen in Figure 4F, we also observed a small but statistically significant increase in Rad52 foci in *sen1-3* mutants compared to wild-type. To determine whether DNA:RNA hybrids contribute to the phenotypes observed in *sen1-3 mrc1Δ*, we repeated the experiments following the overexpression of *RNH1.* This failed to suppress the growth defects and the increase in recombination during S phase (Figures 5F, 5G, and S5B). Similar results were obtained when overexpressing the human ortholog of *RNH1* (Figures S5C and S5D; Wahba et al., 2011; Bonnet et al., 2017).

To analyze whether the increased recombination observed during replication in *sen1-3*, *mrc1Δ sen1-3*, and *hpr1Δ sen1-3* compared to *SEN1* leads to an increase in genomic instability, we measured the rate of direct-repeat recombination using plasmids carrying partially overlapping fragments of the *LEU2* gene separated by 39 or 3,900 nt (plasmids pL and pLYΔNS, respectively) (Mischo et al., 2011; González-Aguilera et al., 2008). We observed, as previously described, that recombination increased with the length of the transcript. Moreover, although *mrc1Δ sen1-3* showed a modest increase in recombination compared to *mrc1Δ* for both plasmids, *hpr1Δ sen1-3* showed greater increases in the rate of recombination (Figure 5H). Furthermore, we tested for defects in mini-chromosome maintenance by transforming a single-copy plasmid carrying an *ADE2* gene and scoring for the rate of plasmid loss in the absence of selective pressure by measuring the rate of white colonies (carrying the plasmid) and red (without the plasmid). Cells carrying the *sen1-3* allele showed higher levels of plasmid loss,

(F) The indicated strains, carrying a *RAD52-GFP* allele with or without the *GAL-RNH1* construct, were grown as shown in Figures S4A–S4D. Samples were taken at the indicated time points, fixed, and analyzed for the presence of Rad52 foci (triplicate biological repeats). n.s., not significant; $^{**}p < 0.05$; $^{***}p < 0.01$.
(G) The indicated strains were grown to exponential phase at 28°C; DNA:RNA hybrids were analyzed by immunofluorescence of chromosome spreads (triplicate biological repeats). Samples were treated in parallel with RNase H.



(legend on next page)

exacerbated in the absence of *MRC1* and *HPR1* (Figures 5I and S5E). Strikingly, *sen1-3 hpr1Δ* completely failed to retain the plasmid. The addition of a second origin of replication did not rescue the chromosome maintenance defects, and overexpression of *RNH1* only partially suppressed the defects in *hpr1Δ sen1-3* cells (Figure S5F).

DISCUSSION

Here, we have shown that Sen1 is a bona fide partner of the yeast replisome. The N-terminal domain mediates binding to replisomes, mainly via Ctf4 and Mrc1. Additional binding partners of Sen1 are likely because Sen1 shows some residual interaction with replisomes in the absence of Ctf4 and Mrc1. It is not yet clear whether multiple Sen1 molecules are recruited to RFs by independently binding separate subunits of the replisome with different affinities or whether multiple replisome components coordinately bind a single Sen1 to increase its strength of interaction. IPs of the N-terminal domain of Sen1 suggest a competition between Mrc1 and Ctf4 for Sen1 as Mrc1 binding increases in *ctf4Δ* cells (Figures 2D and 2E). This supports the multiple independent binding hypothesis. However, deletion of either *CTF4* or *MRC1* decreases overall binding of Sen1 to the replisome (Figures 2F, 2H, S2C, and S2D), compatible with a cooperative recruitment of Sen1. Interestingly, the mutation of three amino acids in *sen1-3* abolishes binding to both Ctf4 and Mrc1. Thus, the mutated residues either correspond to the direct interaction site for both Ctf4 and Mrc1 or they cause a change in conformation of a larger section of Sen1, thus affecting two distinct binding surfaces for Ctf4 and Mrc1. Both hypotheses are compelling, and further work is needed to determine which is correct.

The *sen1-3* allele is a separation of function mutant that breaks the interaction with the replisome without affecting the binding to RNAPII or transcription termination (Figures 3E, 4A, and 4B). However, we cannot exclude that *sen1-3* might affect other Sen1 interactions beyond the replisome. In addition, minimal levels of interaction with replisomes might be retained in *sen1-3* cells, thus weakening the severity of the phenotype observed. Nevertheless, this allele provides us with a tool to dissect the function of the helicase at RFs without affecting its catalytic activity and the bulk of its transcription functions.

It has been previously proposed (using the *sen1-1* allele or Sen1 depletion) that Sen1's presence at RFs is required to quickly remove the R-loops accumulating and interfering with RF progression (Alzu et al., 2012; Brambati et al., 2018; Mischo et al., 2011). In our experimental setting, however, loss of Sen1 from RFs did not show increases in DNA:RNA hybrids or dramatic defects in RF progression (Figures 4G and 5C). In fact, the loss of Sen1 from the replisome only leads to modest defects (small increases in post-replicative recombination and instability of mini-chromosomes; Figures 4F, 5D, and 5I). This suggests that when Sen1 is proficient in transcription termination, there might be enough redundancy at the RFs to deal with DNA:RNA hybrids. However, we observe lethality or severe growth defects when the *sen1-3* allele is present in genetic backgrounds with high endogenous levels of R-loops, such as *rnh1Δ rnh201Δ* and *hpr1Δ.* This supports the idea of an important role for Sen1 in dealing with DNA:RNA hybrids at RFs. Surprisingly, we do not observe an increase in DNA:RNA hybrids levels in *sen1-3* and in *hpr1Δ sen1-3* cells (Figures 4G, S4E, and S4F). Moreover, although increased levels of R-loops have been described for *top1Δ* (El Hage et al., 2010), *pif1Δ* (Boulé and Zakian, 2007; Tran et al., 2017), *sgs1Δ* (Chang et al., 2017), or *mlp1Δ* (García-Benítez et al., 2017), these deletions do not show defects in cell viability or DNA damage sensitivity in combination with *sen1-3* (Figure 5A). This suggests that not all increases in R-loops might be necessarily toxic in *sen1-3* cells. One possibility is that different mutations might lead to dissimilar levels or distinct biochemical features of the R-loops. Moreover, different genetic backgrounds might lead to the accumulation of DNA:RNA hybrids at different sites of the genome (as recently observed; Costantino and Koshland, 2018). Therefore, Sen1 association with the replisome might become critical for the timely resolution of some DNA:RNA hybrids in specific circumstances.

The recruitment of Sen1 at RFs also appears to promote DNA replication independently of R-loops. In fact, in *sen1-3* cells, overexpression of *RNH1* fails to suppress the higher levels of recombination observed in *sen1-3* (Figures 4F and 5D). Given the prominent role of Sen1 in transcription termination described in the literature, it is tempting to speculate that Sen1 might remove transcribing or stalled RNA polymerases at RFs (Han et al., 2017; Porrua and Libri, 2013). Alternatively, Sen1 might

Figure 5. *sen1-3* Presents Synthetic Defects with *mrc1Δ*, *ctf18Δ*, and *rad53Δ*, Leading to Increased Recombination and Mini-chromosome Loss

(A) Summary of the genetic interactions tested with the *sen1-3* allele. Some double mutants (orange line) showed marked differences in temperature sensitivity and DNA damage sensitivity although others did not (green line).

(B) Examples of the defects observed with *sen1-3.* Serial dilution spotting of the indicated strains is shown. The double mutant *rad53Δ sml1Δ* is indicated as *rad53Δ*.

(C) The indicated strains were arrested in G1, shifted to 37°C for 1 h, and released in S phase at 37°C. FACS samples were taken at the indicated times. Red bar, length of DNA replication; green arrow, beginning of the end of mitosis.

(D) Cells, carrying a *RAD52-GFP* allele, were treated as in (C). Samples were taken at the indicated time points, fixed, and analyzed for the presence of Rad52 foci (triplicate biological repeats). $^{**}p < 0.05$; $^{***}p < 0.01$.

(E) Examples of the microscopy data of the experiment in (D). Scale bars represent 5 μm.

(F) Serial dilution spotting of the indicated strains is shown. Cells (+*RNH1*) carry an ectopic *GAL1-RNH1* construct.

(G) *RNH1* overexpression does not suppress the increase in recombination in *mrc1Δ sen1-3* cells. Cell cultures were treated as in (C), except they were grown in YPGAL medium (triplicate biological repeats). $^{**}p < 0.05$; $^{***}p < 0.01$.

(H) The *sen1-3* allele causes an increase in recombination. The cells were transformed with the plasmids pL or pLYΔNS. The ratio of the number of the colonies carrying a recombinant plasmid (*LEU2*) over the total number of cells carrying a plasmid (*URA3*) is shown.

(I) Cells were transformed with plasmids carrying an *ADE2* marker and 1 or 2 origins. Percentage of white colonies over the total number of colonies scored is shown (a measure of genome stability; $^{***}p < 0.5\ 10^{-7}$).

be required to remove other barriers to fork progression, or *RNH1* overexpression might not be sufficient to remove DNA:RNA hybrids present at RF with kinetics similar to Sen1, thus leading to increased fork stalling. In either case, we observe that cells rely on the functions of Mrc1 to promote fork progression and minimize DNA recombination in a *sen1-3* background (Figures 5B–5G). Interestingly, we observe that three key mediators and effectors of the S phase checkpoint (*MRC1*, *CTF18*, and *RAD53*) genetically interact with *sen1-3*. We did not observe any synthetic defects between *sen1-3* and either *mec1Δ sml1Δ*, *tel1Δ*, or *mec1Δ sml1Δ tel1Δ* (not shown). This raises the possibility that Mrc1, Ctf18, and Rad53 might be involved in the response to defects arising in *sen1-3* cells independently of Mec1 and Tel1. Alternatively, the synthetic defects observed are the consequence of other deficiencies in these cells, independent of the S phase checkpoint response. For example, Mrc1 has a key role in RF progression (Yeeles et al., 2017; Hodgson et al., 2007; Duch et al., 2018).

Given that eukaryotic orthologs of Sen1 contain an extended non-catalytic N-terminal sequence (the function of which is still largely unknown), it will be interesting to investigate further whether Senataxin or any of its paralogs (Aquarius, IGHMBP2, RENT1, and ZNFx1) associate with replisomes in higher eukaryotes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Yeast Strains and Growth Conditions
  - Cell cycle experiments
  - Harvesting cells for IP
  - Western Blots
  - IP
  - MS Analysis of IPs
  - Counting of Rad52-foci to assess DNA damage
  - Chromosome spreads and microscopy
  - Quantification of R-loops
  - Reverse transcription followed by quantitative PCR
  - Cross-linking and analysis of cDNA (CRAC)
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.01.087.

ACKNOWLEDGMENTS

We thank Karim Labib, Benoit Palancade, Andres Aguilera, and Nicholas Proudfoot for strains, antibodies, and plasmids. We are grateful to Karim Labib, Jordi Torres-Rosell, Jonathan Millar, and Andrew McAinsh for feedback. The authors acknowledge CAMDU (Computing and Advanced Microscopy Unit) and Media Preparation Facility at the University of Warwick for their assistance in this work, as well as the help of the high-throughput sequencing core facility of I2BC (Centre de Recherche de Gif-sur-Yvette, France). G.D.P. and R.A. were funded by Cancer Research UK Career Development Fellowship C44595/A16326. R.A. has been funded by Chancellor International Scholarship at the University of Warwick. E.C.L. is funded by the BBSRC MIBTP program. This work was also supported by the Centre National de la Recherche Scientifique (CNRS), l'Agence National pour la Recherche (ANR) grants ANR-12-BSV8-0014-01 and ANR-16-CE12-0022-01 to D.L. and the Labex Who Am I? (ANR-11-LABX-0071 et Idex ANR-11-IDEX-0005-02 to D.L. U.A. is supported by a fellowship from the French Ministry of Research.

AUTHOR CONTRIBUTIONS

R.A., D.L., and G.D.P. conceived the study; R.A., E.C.L., U.A., and G.D.P. performed experiments; all authors analyzed experiments; and R.A. and G.D.P. wrote the manuscript. All authors discussed the data and commented and helped improve the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 5, 2019
Revised: September 6, 2019
Accepted: January 24, 2020
Published: February 18, 2020

REFERENCES

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

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

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol. Cell *23*, 841–851.

Bell, S.P., and Labib, K. (2016). Chromosome duplication in Saccharomyces cerevisiae. Genetics *203*, 1027–1067.

Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Géli, V., de Almeida, S.F., and Palancade, B. (2017). Introns protect eukaryotic genomes from transcription-associated genetic instability. Mol. Cell *67*, 608–621.e6.

Boulé, J.B., and Zakian, V.A. (2007). The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. Nucleic Acids Res. *35*, 5809–5818.

Brambati, A., Zardoni, L., Achar, Y.J., Piccini, D., Galanti, L., Colosio, A., Foiani, M., and Liberi, G. (2018). Dormant origins and fork protection mechanisms rescue sister forks arrested by transcription. Nucleic Acids Res. *46*, 1227–1239.

Burgers, P.M.J., and Kunkel, T.A. (2017). Eukaryotic DNA replication fork. Annu. Rev. Biochem. *86*, 417–438.

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

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

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

Chang, E.Y., Novoa, C.A., Aristizabal, M.J., Coulombe, Y., Segovia, R., Chaturvedi, R., Shen, Y., Keong, C., Tam, A.S., Jones, S.J.M., et al. (2017).

RECQ-like helicases Sgs1 and BLM regulate R-loop-associated genome instability. J. Cell Biol. *216*, 3991–4005.

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

Chávez, S., Beilharz, T., Rondón, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T., and Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. EMBO J. *19*, 5824–5834.

Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. (2012). Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in *Saccharomyces cerevisiae*. Eukaryot. Cell *11*, 417–429.

Costantino, L., and Koshland, D. (2018). Genome-wide map of R-loop-induced damage reveals how a subset of R-loops contributes to genomic instability. Mol. Cell *71*, 487–497.e3.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. *7*, e1002329.

Culbertson, M.R., and Leeds, P.F. (2003). Looking at mRNA decay pathways through the window of molecular evolution. Curr. Opin. Genet. Dev. *13*, 207–214.

De Piccoli, G., Katou, Y., Itoh, T., Nakato, R., Shirahige, K., and Labib, K. (2012). Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases. Mol. Cell *45*, 696–704.

Duch, A., Canal, B., Barroso, S.I., García-Rubio, M., Seisenbacher, G., Aguilera, A., de Nadal, E., and Posas, F. (2018). Multiple signaling kinases target Mrc1 to prevent genomic instability triggered by transcription-replication conflicts. Nat. Commun. *9*, 379.

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

El Hage, A., Webb, S., Kerr, A., and Tollervey, D. (2014). Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. PLoS Genet. *10*, e1004716.

Gambus, A., van Deursen, F., Polychronopoulos, D., Foltman, M., Jones, R.C., Edmondson, R.D., Calzada, A., and Labib, K. (2009). A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. EMBO J. *28*, 2992–3004.

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

García-Pichardo, D., Cañas, J.C., García-Rubio, M.L., Gómez-González, B., Rondón, A.G., and Aguilera, A. (2017). Histone mutants separate R loop formation from genome instability induction. Mol. Cell *66*, 597–609.e5.

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

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc. Natl. Acad. Sci. USA *106*, 9613–9618.

Grubb, J., Brown, M.S., and Bishop, D.K. (2015). Surface spreading and immunostaining of yeast chromosomes. J. Vis. Exp., e53081.

Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat. Struct. Mol. Biol. *15*, 786–794.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. Cell *170*, 774–786.e19.

Han, Z., Libri, D., and Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. Nucleic Acids Res. *45*, 1355–1370.

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA polymerase II and Sen1-dependent transcription termination. Mol. Cell *49*, 55–66.

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

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

Hodgson, B., Calzada, A., and Labib, K. (2007). Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S phase. Mol. Biol. Cell *18*, 3894–3902.

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

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., and Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast *21*, 947–962.

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. Trends Biochem. Sci. *36*, 19–29.

Kamimura, Y., Tak, Y.S., Sugino, A., and Araki, H. (2001). Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. EMBO J. *20*, 2097–2107.

Kanemaki, M., and Labib, K. (2006). Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. EMBO J. *25*, 1753–1763.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522.

Kim, T.S., Liu, C.L., Yassour, M., Holik, J., Friedman, N., Buratowski, S., and Rando, O.J. (2010). RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. Genome Biol. *11*, R75.

Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., and Merrikh, H. (2017). Replication-transcription conflicts generate R-loops that orchestrate bacterial stress survival and pathogenesis. Cell *170*, 787–799.e18.

Leonaitė, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O., and Conti, E. (2017). Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. EMBO J. *36*, 1590–1604.

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

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

Martin-Tumasz, S., and Brow, D.A. (2015). Saccharomyces cerevisiae Sen1 helicase domain exhibits 5′- to 3′- helicase activity with a preference for translocation on DNA rather than RNA. J. Biol. Chem. *290*, 22880–22889.

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

Mischo, H.E., Chun, Y., Harlen, K.M., Smalec, B.M., Dhir, S., Churchman, L.S., and Buratowski, S. (2018). Cell-cycle modulation of transcription termination factor Sen1. Mol. Cell *70*, 312–326.e7.

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

Pfeiffer, V., Crittin, J., Grolimund, L., and Lingner, J. (2013). The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. EMBO J. *32*, 2861–2871.

Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nat. Struct. Mol. Biol. *20*, 884–891.

Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R., and Libri, D. (2012). In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. EMBO J. *31*, 3935–3948.

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

Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. Mol. Cell *36*, 88–98.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-wide mapping of yeast RNA polymerase II termination. PLoS Genet. *10*, e1004632.

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

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol. Cell *24*, 735–746.

Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., and Araki, H. (2003). GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. Genes Dev. *17*, 1153–1165.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol. Cell *23*, 853–864.

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

Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F., and Culbertson, M.R. (1997). The yeast *SEN1* gene is required for the processing of diverse RNA classes. Nucleic Acids Res. *25*, 4778–4785.

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

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

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*, 481–489.

Yeeles, J.T.P., Janska, A., Early, A., and Diffley, J.F.X. (2017). How the eukaryotic replisome achieves rapid and efficient DNA replication. Mol. Cell *65*, 105–116.

Yüce, Ö., and West, S.C. (2013). Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. Mol. Cell. Biol. *33*, 406–417.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-mouse-HRP	Cell Signaling Technology	#7076; RRID:AB_330924
Anti-sheep-HRP	Sigma	A3415; RRID:AB_258076
Anti-Cdc45	Labib Lab	N/A
Anti-Csm3	Labib Lab	N/A
Anti-Ctf4	Labib Lab	N/A
Anti-Dpb2	Labib Lab	N/A
Anti-FLAG	Sigma	F3165; RRID:AB_259529
Anti-HA (12CA5)	Sigma	11583816001; RRID:AB_514505
Sheep IgG	Sigma	S1265; RRID:AB_261431
Anti-Mcm3	Labib Lab	N/A
Anti-Mcm4	Labib Lab	N/A
Anti-Mcm5	Labib Lab	N/A
Anti-Mcm6	Labib Lab	N/A
Anti-Mrc1	Labib Lab	N/A
Anti-MYC	Sigma	M4439; RRID:AB_439694
Anti-Nrd1	Libri Lab	N/A
Anti Nab3	Libri Lab	N/A
Anti-Psf1	Labib Lab	N/A
Anti-Pob3	Labib Lab	N/A
Anti-Pol1	Labib Lab	N/A
Anti-Pol2	De Piccoli Lab	N/A
Anti-Rad53	Abcam	ab166859; RRID:AB_2801547
Anti-Rpo21	Novus Biologicals	NB200-598SS; RRID:AB_2252678
Anti-Sld5	K. Labib	N/A
Anti-TAP-HRP	Sigma	P1291; RRID:AB_1079562
Anti DNA:RNA hybrids S9.6	Kerafast	ENH001; RRID:AB_2687463
Cy3-conjucated anti-mouse	Jackson laboratories	#115165003; RRID:AB_2338680
Anti-dsDNA	Abcam	ab27156; RRID:AB_470907
Chemicals, Peptides, and Recombinant Proteins		
α-factor	Pepceuticals	N/A
AcTEV protease	Thermo-Fischer	12575015
Calmodulin	Sigma	A6112
Complete protease inhibitor cocktail	Roche	11 836 153 001
Dithiothreitol	Sigma	D0632
Dynabeads	Invitrogen	14302D
Ethidium bromide	Sigma	E1510
Hydroxyurea	Sigma	H8627
Methyl methanesulfonate	Sigma	129925
Propidium iodide	Sigma	P4864
Protease Inhibitor Cocktail	Sigma	P8215
Sodium fluoride	Thermo-Fischer	S299500
Zymolyase	Zymo research	#E1005
RNase H	Invitrogen	#18021071

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sodium glycerophosphate	Johnson Matthey	170096
Universal Nuclease	Pierce	88700
Critical Commercial Assays		
Amersham ECL Western Blotting Detection Reagent	GE Healthcare	RPN2108
LightCycler® FastStart DNA Master SYBR Green I	Roche	03003230001
MLV-Reverse Transcriptase	Thermofischer	28025013
QuikChange Lightning Site-Directed Mutagenesis Kit	QIAGEN	#210519
Experimental Models: Organisms/Strains		
S. cerevisiae (from W303) CS1*MATa*	Rothstein's lab	N/A
S. cerevisiae (from W303) CS74*MATa pep4Δ::ADE2*	Lab strain	N/A
S. cerevisiae (from W303) CS1125*MATa TAP-SLD5 (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2*	This study	N/A
S. cerevisiae (from W303) CS1126*MATa SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2*	This study	N/A
S. cerevisiae (from W303) CS1187*MATa TAP-SLD5 (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2 ctf4Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS1353*MATa SEN1-TAP (kanMX) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1416*MATa TAP-MCM3 (kanMX) SEN1-9MYC (hphNT) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1711*MATa TAP-MCM3 (kanMX) GAL1-3HA-ø (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1714*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (2-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1534*MATa TAP-SLD5 (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2 mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS1852*MATa leu2-3,112::GAL1-TAP-ø (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1933*MATa leu2-3,112::GAL1-TAP-SEN1 (1095-2231) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1941*MATa leu2-3,112::GAL1-TAP-SEN1 (2-2231) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1942*MATa leu2-3,112::GAL1-TAP-SEN1 (2-1901) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1943*MATa leu2-3,112::GAL1-TAP-SEN1 (931-2231) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1956*MATa leu2-3,112::GAL1-TAP-SEN1 (2-1103) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1957*MATa leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2030*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (2-622) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2032*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (410-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2056*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-ø (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2058*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2061*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (2-1901) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2062*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (1095-2231) (LEU2)*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. cerevisiae (from W303) CS2148*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (501-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2150*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (622-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2184*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (2-1103) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2188*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (2-2231) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2451*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (931-2231) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2458*MATa/MATα SEN1/SEN1 (931-2231) (HISMX)*	This study	N/A
S. cerevisiae (from W303) CS2582*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (931-2231) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2584*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2603*MATa leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ::ADE2 ctf4Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS2607*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) W773A E774A W777A (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2609*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2611*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) V858A R859A I862A (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2615*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) D876G D877G V880G (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2617*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) V746G D747G P748G I749G (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2623*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660 (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2636*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660A (LEU2) NRD1- 9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2638*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2- 2231)W773A E774A W777A (LEU2) NRD1-9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2640*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2- 2231) D850A E851G V852A L853G L854A (LEU2) NRD1- 9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2642*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2- 2231) V746G D747G P748G I749G (LEU2) NRD1-9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2669*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) (LEU2) NRD1-9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2670*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2- 2231) (LEU2) NRD1-9MYC (HIS3MX) pep4Δ:: ADE2*	This study	N/A
S. cerevisiae (from W303) CS2716*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) D876G D877G V880G (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2718*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) T782G I783G Y784G (LEU2)*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. cerevisiae (from W303) CS2734*MATa rnh1Δ:: hphNT rnh201Δ::HISMX*	Lab strain	N/A
S. cerevisiae (from W303) CS2735*MATα rnh1Δ:: hphNT rnh201Δ::HISMX*	Lab strain	N/A
S. cerevisiae (from W303) CS2791*MATa td-sld3-7 (kanMX) GAL1-UBR1 (HIS3MX) leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2+) pep4Δ:: ADE2*	This study	N/A
S. cerevisiae (from W303) CS2808*MATa SEN1-TAP (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2810*MATa sen1-3-TAP (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2853*MATa SEN1-TAP (kanMX) pep4Δ:: ADE2*	This study	N/A
S. cerevisiae (from W303) CS2854*MATa sen1-3-TAP (kanMX) pep4Δ:: ADE2*	This study	N/A
S. cerevisiae (from W303) CS2859*MATa SEN1-TAP (kanMX) pep4Δ:: URA3 mrc1Δ::hphNT*	This study	N/A
C *S. cerevisiae* (from W303) S2861*MATa sen1-3-TAP (kanMX) pep4Δ:: URA3 mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS2903*MATa td-sld3-7 (kanMX) GAL1-UBR1 (HIS3MX) leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ:: ADE2 ctf4Δ:: kanMX*	This study	N/A
S. cerevisiae (from W303) CS2938*MATα SEN1-TAP (kanMX) hpr1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS2941*MATα sen1-3-TAP (kanMX) hpr1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS2945*MATa SEN1-TAP (kanMX) sml1Δ::HISMX rad53Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2947*MATa sen1-3-TAP (kanMX) sml1Δ::HISMX rad53Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2955*MATa SEN1-TAP (kanMX) ctf18Δ::klTRP1*	This study	N/A
S. cerevisiae (from W303) CS2957*MATa sen1-3-TAP (kanMX) ctf18Δ::klTRP1*	This study	N/A
S. cerevisiae (from W303) CS3167*MATa leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ::ADE2 psf1-1 (ts)*	This study	N/A
S. cerevisiae (from W303) CS3186*MATa leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ::ADE2 mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS3321*MATα leu2-3,112::GAL1-RNH1 (2-348) (LEU2) SEN1-TAP (kanMX) mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS3322*MATa leu2-3,112::GAL1-RNH1 (2-348) (LEU2) sen1-3-TAP (kanMX) mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS3499*MATa SEN1-TAP (kanMX) pep4Δ::ADE2 ctf4Δ::kanMX mrc1-3IAA (HISMX) ADH1-OsTIR1 (klTRP1,URA3)*	This study	N/A
S. cerevisiae (from W303) CS3545*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-sen1-3 (2-2231) (LEU2) rnh1Δ:: hphNT rnh201Δ::HISMX*	This study	N/A
S. cerevisiae (from W303) CS3547*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) (LEU2) rnh1Δ:: hphNT rnh201Δ::HISMX*	This study	N/A
S. cerevisiae (from W303) CS3562*MATa SEN1-TAP (kanMX) pep4Δ::ADE2 ctf4Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS3662*MATa SEN1-TAP (kanMX) mrc1Δ::hphNT leu2-3,112::GAL1-RNH1 (2-348) (LEU2+)*	This study	N/A
S. cerevisiae (from W303) CS3664*MATa sen1-3-TAP (kanMX) mrc1Δ::hphNT leu2-3,112::GAL1-RNH1 (2-348) (LEU2)*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. cerevisiae (from W303) CS3702*MATa TAP-SLD5 (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2 ctf4Δ::kanMX mrc1-3IAA (HISMX) ADH1-OsTIR1 (klTRP1,URA3)*	This study	N/A
S. cerevisiae (from W303) CS3731*MATα SEN1-TAP (kanMX) leu2-3,112::GAL1-RNH1 (2-348) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS3733*MATa sen1-3-TAP (kanMX) leu2-3,112::GAL1-RNH1 (2-348) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS3796*MATa SEN1-TAP (kanMX) mad2Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS3797*MATα sen1-3-TAP (kanMX) mad2Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS3903*MATa leu2-3,112::GAL1-RNH1 (2-348) (LEU2) SEN1-TAP (kanMX) hpr1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS3905*MATa leu2-3,112::GAL1-RNH1 (2-348) (LEU2) sen1-3-TAP (kanMX) hpr1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS4296*MATa SEN1-TAP (kanMX) chl1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS4298*MATα sen1-3-TAP (kanMX) chl1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS4312*MATa NRD1-TAP (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3-CP ADE2*	This study	N/A
S. cerevisiae (from W303) CS4314*MATa SEN1-TAP (kanMX) pep4Δ::ADE2 rpb1-1 (ts)*	This study	N/A
S. cerevisiae (from W303) DLY2057*MATa sen1-1 (ts)*	Lab strain	N/A
S. cerevisiae (from W303) DLY2281*MATa upf1Δ::TAP::klTRP1*	Lab strain	N/A
S. cerevisiae (from W303) DLY3111*MATa sen1-1 (ts) upf1Δ::TAP::klTRP1*	This study	N/A
S. cerevisiae (from W303) DLY3190*MATa SEN1-TAP (kanMX) upf1Δ::TAP::klTRP1*	This study	N/A
S. cerevisiae (from W303) DLY3191*MATa sen1-3-TAP (kanMX) upf1Δ::TAP::klTRP1*	This study	N/A
Oligonucleotides		
DL377ATGTTCCCAGGTATTGCCGA	This study	N/A
DL378ACACTTGTGGTGAACGATAG	This study	N/A
DL474GCAAAGATCTGTATGAAAGG	This study	N/A
DL475CGCAGAGTTCTTACCAAACG	This study	N/A
DL481TAAATGGCCAACCGCTGTTG	This study	N/A
DL482CCAGCGTACTGCACGCCAGG	This study	N/A
DL1119AAGTGACGAAGTTCATGCTA	This study	N/A
DL1120TCCGTGTCTCTTGTCCTGCA	This study	N/A
Recombinant DNA		
pYM-N24	Janke et al., 2004	Euroscarf
pCS14pRS305-*GAL1-TAP-Ø*	This study	N/A
pCS25pRS305-*GAL1-3HA-Ø*	This study	N/A
pCS26pRS305-*GAL1-3HA-SEN1 (2-931)*	This study	N/A
pCS30pRS305-*GAL1-TAP-SEN1 (2-931)*	This study	N/A
pCS31pRS305-*GAL1-TAP-SEN1 (2-1103)*	This study	N/A
pCS32pRS305-*GAL1-TAP-SEN1 (931-2231)*	This study	N/A
pCS33pRS305-*GAL1-TAP-SEN1 (1095-2231)*	This study	N/A
pCS39pRS305-*GAL1-TAP-SEN1 (2-2231)*	This study	N/A
pCS40pRS305-*GAL1-TAP-SEN1 (2-1901)*	This study	N/A
pCS42pRS305-*GAL1-3HA-SEN1 (2-622)*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pCS43pRS305-*GAL1-3HA-SEN1 (410-931)*	This study	N/A
pCS59pRS305-*GAL1-3HA-SEN1 (501-931)*	This study	N/A
pCS61pRS305-*GAL1-3HA-SEN1 (622-931)*	This study	N/A
pCS118pRS305-*ACT1-3HA-SEN1 (931-2231)*	This study	N/A
pCS120pRS305-*ACT1-3HA-SEN1 (2-2231)*	This study	N/A
pCS123pRS305-*ACT1-3HA-SEN1 (2-2231) W773A E774A W777A*	This study	N/A
pCS124pRS305-*ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660A*	This study	N/A
pCS125pRS305-*ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A*	This study	N/A
pCS127pRS305-*ACT1-3HA-SEN1 (2-2231) D876G D877G V880G*	This study	N/A
pCS128pRS305-*ACT1-3HA-SEN1 (2-2231) V746G D747G P748G I749G*	This study	N/A
pCS129pRS305-*ACT1-3HA-SEN1 (2-2231) T782G I783G Y784G*	This study	N/A
pCS188pRS305-*GAL1-RNH1 (2-348)*	This study	N/A
pCS196pRS424-*GPD-hsRNASEH1* (2-286)	From Palancade's lab	N/A
pCS197pRS315-*ADE2*	This study	N/A
pCS198pRS315-*ADE2-ARS306*	This study	N/A
pLpRS316-*leu2Δ3′-39bp-leu2Δ5′*	From Aguilera's lab	N/A
pLYΔNSpRS316-*leu2Δ3′-3900bp-leu2Δ5′*	From Aguilera's lab	N/A
Software and Algorithms		
Excel	Microsoft	RRID:SCR_016137
Illustrator	Adobe	RRID:SCR_014198
ImageJ	NIH	https://imagej.nih.gov/ij/; RRID:SCR_003070
Photoshop	Adobe	RRID:SCR_014199
PredictProtein.org		https://www.predictprotein.org/
RStudio	RStudio	RRID:SCR_000432

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed and will be fulfilled by the Lead Contact, Dr Giacomo De Piccoli (g.de-piccoli@warwick.ac.uk). All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Saccharomyces cerevisiae is the experimental model used in this study. All strains are isogenic to W303, and are listed in the Key Resources Table.

METHOD DETAILS

Yeast Strains and Growth Conditions

All yeasts were grown in YP medium supplemented with either glucose (YPD) or galactose (YPGAL) or raffinose (YPRAF) to a final concentration of 2% (w/v). For solid media, the same formulation was used, but with a final concentration of 1% (w/v) agar. Yeasts were grown at 24, 28, 30 and 37°C, depending on their viability at the different temperatures and as required by the experimental design. For all experiments, the control and test strains were subjected to the same conditions, including temperature.

For cell spotting experiments, cells were grown on non-selective media until colonies were judged to be sufficiently big. Five discrete colonies from individual strains were added to sterile deionised water to create a cell suspension. From this suspension, serial dilutions (0.5 x10^6, 0.5 x10^5, 0.5 x10^4 and 0.5 x10^3 cells/ml) were generated. 10 μL of each suspension was pipetted onto the appropriate media and grown for up to 5 days at the required temperatures.

To assess the genetic interaction between two or three genes, parents carrying the appropriate alleles were first crossed. Analysis of the meiotic progeny was conducted by inducing sporulation of the diploid strains in sporulation medium for 3-5 days at 24°C. Asci

were treated with a 1:10 dilution of β-glucoronidase from *Helix pomatia* (Sigma) for 30 minutes, followed by tetrad dissection onto a YPD plate using a Singer MSM400 micromanipulator. Plates were incubated for 3-4 days at the appropriate temperature.

For the plasmid recombination assay, eight independent clones carrying the appropriate plasmid (pL or pLYΔNS) were each plated in medium lacking leucine (to select for recombination) or lacking uracil (marker for the presence of the plasmid) at 24°C. The experiment was repeated in triplicate. For plasmid loss assays, cells were transformed with the required plasmid (pCS197 or pCS198) and plated on minimum medium lacking leucine and incubated at 24°C. Colonies were left to grow until single isolated colonies were sufficiently big. Five to seven colonies for each strain were then picked, resuspended in sterile water and counted. Around 200-150 cells were then plated onto YPD and incubated at 24°C until red/white coloring was clearly visible. Cells were then incubated at 4°C for three days. We considered white and sectored colonies as white while only fully red colonies were scored as red. The experiment was repeated twice. The plasmid loss assay with or without *GAL-RNH1* was conducted in a similar manner, except that cells were grown and transformed in medium containing galactose and selected in medium lacking adenine (*LEU2* is the reporter gene for the *GAL1-RNH1* construct). Colonies were grown for longer periods of time before colonies were sufficient size big and were plated onto non-selective medium containing galactose.

Cell cycle experiments

Cells were diluted from an inoculum to a density of 0.35×10^7 cells/ml in a suitable volume and left to grow to a final density of 0.7×10^7 cells/ml. The cells were then synchronized in G1 by adding α-factor to a final concentration of 7.5 μg/ml. After the first 90 min, α-factor was added every 30 min to a 3.25 μg/ml final concentration to maintain the cells in G1. When the cultures were shifted to 37°C, cells were spun down and resuspended in pre-warmed medium containing 7.5 μg/ml α-factor. Cells were released from the arrest by washing the cells twice with medium without α-factor. In all experiments in which cells were collected for IPs, cells were grown at 24°C and released into S phase for 30 min, unless stated otherwise in the figure legend. For expressing constructs under the *GAL1* promoter, strains were grown in YPRAF, arrested in G1 using α-factor, upon which YPGAL was substituted for YPRAF. Alternatively, YPGAL was used throughout the experiment (appropriate for constructs that were labile).

Harvesting cells for IP

Harvested cells were immediately cooled to 4°C by washing with an ice-cold solution of HEPES-KOH (pH 7.9), followed by a wash in a solution of 100 mM HEPES-KOH (pH 7.9), 50 mM potassium acetate, 10 mM magnesium acetate and 2 mM EDTA-KOH, still at 4°C. After the wash, the solution was discarded and the cells were re-suspended in a fresh quantity of the same solution supplemented with protease and phosphatase inhibitors, so that the ratio of wet mass of the cells to the final mass of the suspension was either 1:4 (for 250 mL cultures) or 4:5 (for 1 l cultures). The re-suspended cells were immediately flash-frozen by pipetting into a flask holding liquid nitrogen. The frozen cells were kept at −80°C until use for IP. Before freezing, some cells were fixed in 70% (v/v) ethanol to test that cells did not progress through the cell cycle during sample preparation.

For cells with inducible constructs, cultures were grown as described above in YPRAF. After the cells were arrested in G1, the culture was substituted with YPGAL (supplemented with α-factor) to induce transcription from the *GAL1* promoter. Harvesting of G1 cultures can be performed prior to or after induction according to the experimental setup. After 35 min or 1 h of induction, the cells were released in S phase as described above and harvested either 30 min (24°C) or 20 min (30°C or 37°C) post-release. For temperature-sensitive strains or strains tagged with a temperature-degron (e.g *psf1-1*, *td-sld3-7* and *rnh1Δ rnh201Δ ACT1-sen1-3*), the strains were grown and synchronized in G1 at 24°C as described above. Once synchronized and, (optionally) constructs transcriptionally induced, the cells were shifted to 37°C for 1 h. α-factor was added every 20 min to maintain the cells in G1 to a final concentration of 7.5 μg/ml for 1 h. Synchronicity was monitored visually using a microscope and by harvesting a 1 mL sample of the culture by fixing in 70% (v/v) ethanol for flow-cytometric analysis. The cells were then washed and released in S phase. The cells were harvested 20 min after release, including for the *psf1-1* strains that do not actually undergo DNA replication at 37°C as the GINS complex is destabilized. For crosslinking IPs, cells cultures were incubated with formaldehyde for 25 min and treated as in De Piccoli et al. (2012).

Western Blots

Protein samples (TCA-precipitated and non-treated cell extracts, as well as IPs) were run on 5, 6, 7, 8 or 10% polyacrylamide gels. The protein bands were then transferred onto nitrocellulose or PVDF membranes. The proteins bands were then probed with the appropriate primary antibodies for 1 h in a solution of 5% (w/v) skimmed milk in TBST, washed thrice for 5 min in fresh TBST, probed with the appropriate HRP-bound secondary antibody (if any, refer to Key Resources Table) and washed thrice again for 5 min in fresh TBST. The membrane was then treated with the western blotting reagents and the resulting chemiluminescent signal was captured using either films or a digital camera (G:BOX, Chemi XRG, Syngene).

IP

IPs were conducted as previously described (De Piccoli et al., 2012). In brief, cells previously harvested were lysed using a mechanised pestle and mortar at −80°C (Spex Sample Prep, 6870). 1 g of lysate is considered equivalent to 1 ml. To 1 volume of thawed lysate, ¼ volume of a solution of 50% (v/v) glycerol, 100 mM HEPES-KOH (pH 7.9), 50 mM potassium acetate, 50 mM magnesium acetate, 0.5% (v/v) Igepal® CA-630, 2mM EDTA supplemented with protease and phosphatase inhibitors was added. Pierce

Universal Nuclease was added to a final concentration of 0.4 U/μl and samples were left on a rotating platform at 4°C for 30 min. After incubation, the sample was clarified by stepwise centrifugation at 18,700*g* and then at 126,600*g*. The supernatant was isolated, 50 μL of which was added to 100 μL of 1.5 x Laemmli buffer (cell extract). The remaining cell extract was incubated with 100 μL of TAP-beads for 2 h (M-270 Dynabeads® Epoxy beads bound to an anti-sheep IgG). Beads were washed with solutions of 100 mM HEPES-KOH (pH 7.9), 50 mM potassium acetate, 50 mM magnesium acetate, 2 mM EDTA, 0.1% (v/v) Igepal® CA-630 thrice. After washing, 100 μL of 1 x Laemmli buffer was added to the 100 μL of TAP-beads and boiled for 4 min. Crosslinking IPs were conducted as in De Piccoli et al. (2012).

When scaling up was necessary (using 1 l of cells instead of 250 ml), a few changes were implemented to the protocol. Notably, the concentration of the Pierce Universal Nuclease was increased four-fold to a final concentration of 1.6 U/μl and incubation with the nuclease was increased from 30 to 40 min.

MS Analysis of IPs

The samples were processed as above. Following the washes, TAP-Sen1 (2-931) protein was released by using the AcTEV® protease at 24°C for 2 h. Thereafter, the resultant CBP-Sen1 (2-931) (CBP: calmodulin-binding protein) and its specific interactors were incubated with pre-washed calmodulin beads at 4°C for 2 h. After washing, 30 μL of 1 X Laemmli was added to the calmodulin beads and boiled for 4 min. The samples from the four biological replicates were pooled together, flash-frozen on dry ice and stored at −80°C. The samples were then run on commercially sourced 4%–12% acrylamide gel for a short distance (~1 cm). The gel was then cut in thin slices and processed and analyzed by MS Bioworks, USA.

Counting of Rad52-foci to assess DNA damage

Cells carrying the *RAD52-GFP* allele were first grown in liquid medium and synchronized in G1. Cells were released and harvested at different times after release, corresponding to different phases of the cell cycle. Paraformaldehyde was added to the cell suspensions to a final concentration of 3% (w/v) and the samples were incubated at room temperature for 10 min. The cells were then washed with PBS at room temperature. Finally, the samples were re-suspended in fresh PBS and kept at 4°C overnight.

Less than 24 h after fixation (to minimize signal lost due to alteration of the GFP protein), the samples were re-suspended in 500 μL of fresh PBS to which the DNA stain DAPI was added to a final concentration of 1 μg/ml. The samples were incubated at room temperature for 10 min to allow for staining of the DNA. The cells were then washed with PBS to improve the signal-to-noise ratio of the DAPI staining. The cells were then brought to a suitable dilution prior to pipetting on a glass slide onto which a coverslip is applied. Images of cells were acquired (brightfield, ~510 nm emission (GFP), ~460 nm (DAPI)) using a Personal DeltaVision (Applied Precision). The images were analyzed using ImageJ and the number of Rad52-foci were counted. An average of three experiments is shown in the figures.

Chromosome spreads and microscopy

Chromosome spreads were performed as previously described (Wahba et al., 2011; Grubb et al., 2015). Exponentially growing asynchronous cultures were grown in YPD at 28°C. $2x10^8$ cells were harvested and spheroplasted (0.1M potassium phosphate (pH 7.4), 1.2 M sorbitol, 0.5 mM MgCl2, 40 mM DTT, 20 U zymolyase at 30°C for 1 h or until > 90% of cells lysed following addition of 2% sarcosyl. Cells were then washed and resuspended in ice cold 1 M sorbitol (pH 6.4), 0.1 M MES, 0.5 mM MgCl2, 1 mM EDTA to stop spheroplasting reaction. 20 μL of cell suspension was placed onto a slide, followed by 40 μL of fixative (4% paraformaldehyde (w/v), 3.4% sucrose (w/v)), then lysed using 80 μL of 1% lipsol (v/v) for 2 min, followed by addition of 80 μL of fixative and spread across the surface of the slide to dry overnight. Slides pre-treated with RNase H were incubated with 4U of RNase H diluted in 400 μL of 5mg/ml BSA for 1 h at 37°C prior to immunostaining. Slides were immunostained for DNA:RNA hybrids using mouse monoclonal antibody S.96 (Kerafast) diluted 1:2000 (0.25 μg/ml) in blocking buffer (5% BSA, 0.2% milk, 1XPBS) for 1 h. The secondary antibody, Cy3-conjucated goat anti-mouse (Jackson laboratories) was diluted 1:2000 in blocking buffer and incubated in the dark for 1 h. Indirect immunofluorescence was observed using a Deltavision 1 microscope with a 100 × /NA 1.4 objective. Image analysis was performed using ImageJ and Adobe Photoshop. An average of three experiments is shown in the figures.

Quantification of R-loops

Cells growing in liquid culture was harvested and re-suspended in lysis solution (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 3% (w/v) SDS). To a volume of cell suspension, an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) (Acros Organics) and another volume of nuclease-free deionised water were added. The cells were then lysed mechanically using glass beads and DNA was isolated by incubating the soluble cell extract to ethanol to a final concentration of 70% (v/v). The DNA was washed with fresh ethanol and re-suspended in nuclease-free TE supplemented with 50 μg/ml RNase A and incubated at 37°C for 1 h only.

The concentration of genomic material was estimated by measuring absorbance at 260 nm. For each sample, 1 μg/μl, 0.5 μg/μl and 0.25 μg/μl dilutions of DNA was prepared, using nuclease-free water. 2 μL of each dilution was treated with either 1U of RNaseH (Invitrogen, #18021071) or 1 U of RNaseH and 1 U of RNase III (Invitrogen, #AM2290) with similar results at 37°C for 1h. As a control, untreated samples were also incubated at 37°C for 1 h. The remaining DNA was then added to 200 μL of 2 X SSC hybridization buffer (0.3M NaCl, 30mM trisodium Citrate, pH 7.0) and transferred to a pre-equilibrated hybond-N+ nylon membrane (GE healthcare, #RPN203B) under vacuum. The DNA was cross-linked to the membrane using UV prior to blocking in either 5% (w/v) milk (anti-R

loops) or 5% (w/v) BSA (anti ds-DNA) at 24°C for 1 h. The membranes were then incubated overnight in 5% milk supplemented the primary antibody at 4°C overnight. After thrice washing in TBST for 30 min, the membrane was incubated with anti-mouse IgG-HRP at 24°C for 1 h. The membranes were treated with ECL, and chemiluminescent signal was visualized using a camera (G:BOX, Chemi XRG, Syngene).

Reverse transcription followed by quantitative PCR

Cells were grown to exponential phase and incubated at permissive (24°C) or non-permissive temperatures (37°C) for 3 h to induce the *sen1-1* phenotype before collection. Analysis was performed in parallel in an *upf1Δ* background for detecting elongated RNA species derived from termination failure that might be degraded in the cytoplasm. The ratio of the read-through fraction over the total amount of *SNR13* RNA is shown as a proxy of transcription termination levels. The mean of three experiments is shown. Error bars represent the standard deviation. Cells were grown in logarithmic phase, and 6 OD_{600} worth of cells were pelleted. Total RNAs were extracted by resuspending cell pellets in 1 volume of acidic phenol (pH 4.3) supplemented with 1 volume of AES Buffer (50 mM Sodium Acetate pH 5.5, 10 mM EDTA, 1% SDS). Mixtures were incubated at 70°C with agitation (1,400 rpm) for 30 min in a thermomixer (Eppendorf), before being centrifuged at *20,000 g* at 4°C for 10 min. Aqueous phases were recovered and subjected to one extra round of hot acidic phenol extraction, followed by one round of chloroform extraction. Total RNAs were finally precipitated with absolute ethanol and sodium acetate pH 5.5, washed once with 70% Ethanol, dried on a SpeedVac (Thermo) and resuspended in 30 μL of RNase-free H_2O. 60-120 μg of total RNAs were recovered routinely.

Reverse transcription was performed using random hexamer-primers annealing at multiple *loci* in the *S. cerevisiae* genome and with oligos dT. 4 μg of total RNAs were mixed to 200 ng of random hexamers and 0.5 μM of oligos dT in a 20 μL reaction containing 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl2 and 5 mM DTT. Samples were first incubated for 15 min at 70°C to allow RNA denaturation. Then temperature was slowly decreased to 37°C to allow annealing of primers. Lastly, synthesis of cDNAs was performed by adding 200 units of MLV-reverse transcriptase for 45 min at 37°C.

To assess the amount of cDNAs reverse transcribed, quantitative PCR (qPCR) was carried out using two different primer pairs for each target (*SNR13, NEL250c, ACT1*). These allowed the amplification of a product covering either ~300 bp of the 3′ end of *ACT1* (DL377/DL378 primer pair) or ~70 bp in the read-through region of *SNR13* (DL1119/DL1120 primer pair) or ~140 bp in the body of *NEL025c* (DL474/DL475 primer pair) or ~70 bp in the read-though region of *NEL025c* (DL481/DL482 primer pair). qPCR was performed in a 10 μL reaction by mixing 2 μL of the reverse transcribed cDNAs to 5 μL of LightCycler® 480 SYBR Green I Master and 2.5 pmol of both the forward and the reverse primer.

Cross-linking and analysis of cDNA (CRAC)

The CRAC protocol used in this study is derived from Granneman et al. (2009) with a few modifications as described in Candelli et al. (2018). Raw data processing has been performed as described in Candelli et al. (2018). Metagene analysis has been performed as follows: for the CUTs presented in Table S1, we retrieved the polymerase reads count at every position around the features (3′ or 5′ end) and plotted the mean over all the values for these positions in the final aggregate plot. Analysis has been performed in the R Studio environment.

QUANTIFICATION AND STATISTICAL ANALYSIS

Where applicable, data was presented as the average ± standard deviation. t tests were used to compare population means. Statistically significant differences were indicated as such by indicating the value range of the p values.

DATA AND CODE AVAILABILITY

The published article includes all datasets generated or analyzed during this study. The raw data of the metagene analysis of the CUTs shown in Figure 4B are included in Table S1. This study did not generate any unique code.

Cell Reports, Volume *30*

Supplemental Information

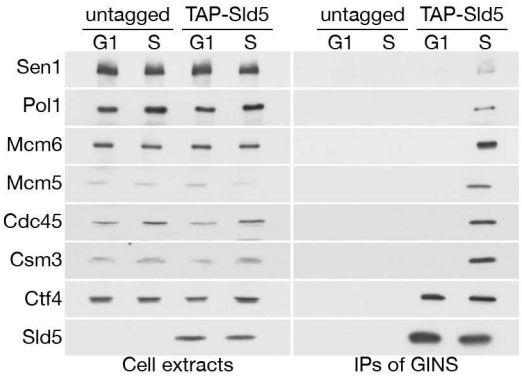
Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability

Rowin Appanah, Emma Claire Lones, Umberto Aiello, Domenico Libri, and Giacomo De Piccoli

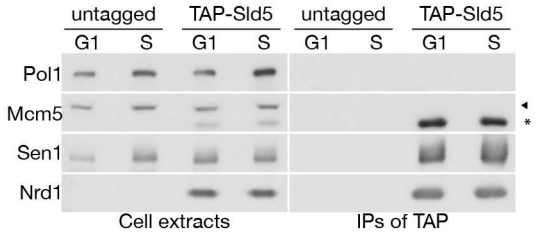
A)

Protein ID	MW (kDa)	TAGGED (peptide N)	UNTAGGED (peptide N)
CTF4	104	365	0
MCM4	115	123	0
POL2	256	97	0
SEN1	253	96	0
TOP1	90	85	0
DPB3	59	83	17
MCM3	108	61	0
SLD5	40	34	3
MCM6	113	34	0
DPB2	61	27	6
MRC1	124	25	0
MCM5	86	24	0
HHF1	11	22	0
TOF1	141	22	0

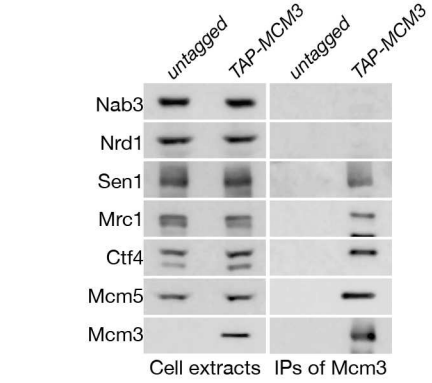
B)



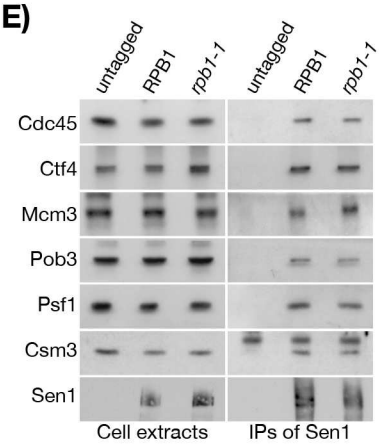
C)



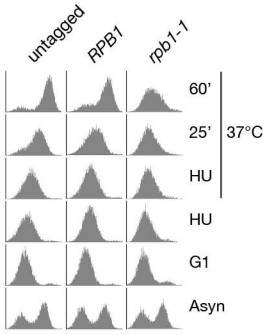
D)



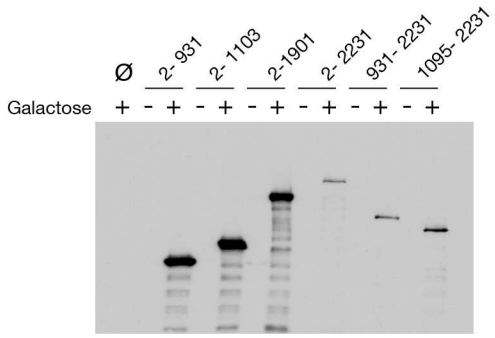
E)



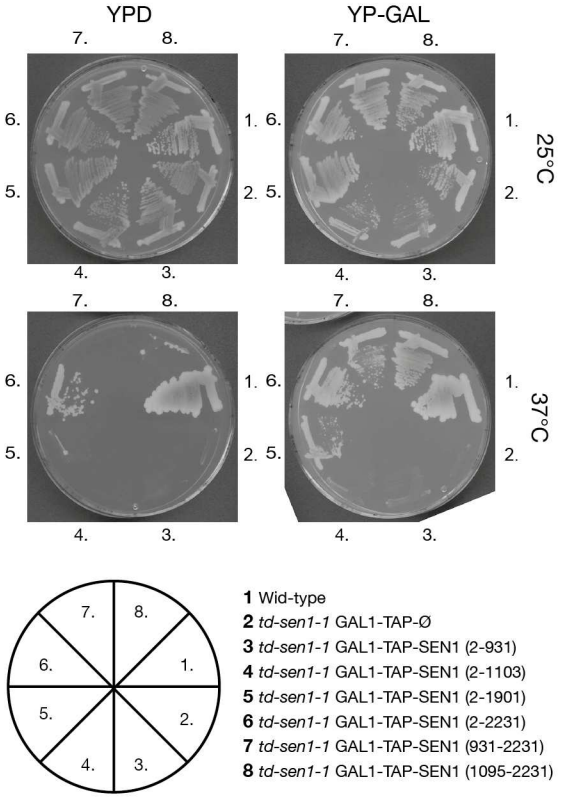
F)



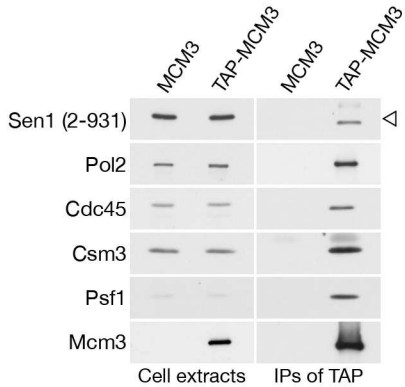
G)



H)



I)



J)

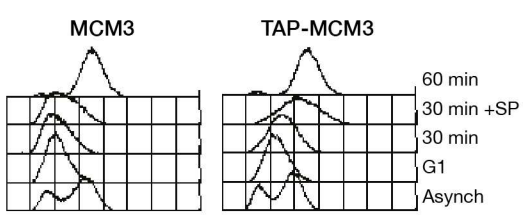


Figure S1. The replisome binds Sen1 in S phase (related to Fig 1)**. A)** Example of the mass spectrometry analysis obtained from the double purification of Sld5 and Mcm4. **B)** Cells carrying the *SEN1-9MYC* allele with a *SLD5* or *TAP-SLD5* allele were synchronously released from G1 into S phase for 30 min at 24°C. Cell extracts were incubated with anti-TAP beads and analysed by immunoblotting. **C)** Nrd1 does not interact with the replisome. *NRD1* or *NRD1-TAP* cells were released from G1 arrest into S phase for 30 min at 24°C. Cell extracts were incubated with anti-TAP beads and analysed by immunoblotting. **D)** Mcm3 immunoprecipitates Sen1 but neither Nrd1 nor Nab3. *MCM3* or *TAP-MCM3* cells were arrested in G1 and released into S phase for 30 min at 24˚C. Cell extracts were incubated with anti-TAP beads and analysed by immunoblotting. **E)** Sen1 interacts *in vivo* with the replisome independently of RNAPII transcription. Wild type or *rpb1-1* cells, either carrying an untagged or TAP-tagged allele of *SEN1*, were arrested in G1 and released in medium containing 0.2 M HU for 75 min at 24˚C. Cultures were then shifted to 37˚C for 1 h. Inactivation of *rpb1-1* cells at 37˚C for 1 h has been shown to lead to a substantial loss of Rpb1-1 from chromatin (Zanton and Pugh, 2006; Kim *et al.*, 2010), to a loss of elongation factors Spt5 and Spt16 (Tardiff, Abruzzi and Rosbash, 2007), to a loss of Sen1 recruitment at highly transcribed genes (Alzu *et al.*, 2012) and to the termination of transcription (Nonet *et al.*, 1987). Cells were then released for 25 min at 37˚C in fresh medium so to allow the synthesis of the bulk of the DNA. Cell extracts were incubated with anti-TAP beads and analysed by immunoblotting. **F)** FACS analysis of the experiment in **E)**. **G)** Cells carrying several different N-terminally tagged truncations of *SEN1* under the *GAL1* promoter were grown to exponential phase in YPRAF, divided in two cultures and transferred to either fresh YPRAF or to YPGAL for 2 h. Protein extracts were analysed by immunoblotting with an anti-HA antibody. **H)** (Top) Cells carrying a temperature-sensitive allele *td-sen1-1* and different fragments of *SEN1* under the *GAL1-3HA* promoter were plated, according to the schematic presented, on YPD or YPGAL and incubated at either 24°C or 37°C. (Bottom) schematic of the plated strains. **I)** The interaction of Sen1 (2-931) with TAP-Mcm3 during S phase is specific. Experiments carrying an untagged or a TAP-tagged allele of *MCM3* were conducted as in Fig 1C. **J)** FACS samples from the experiment are shown.

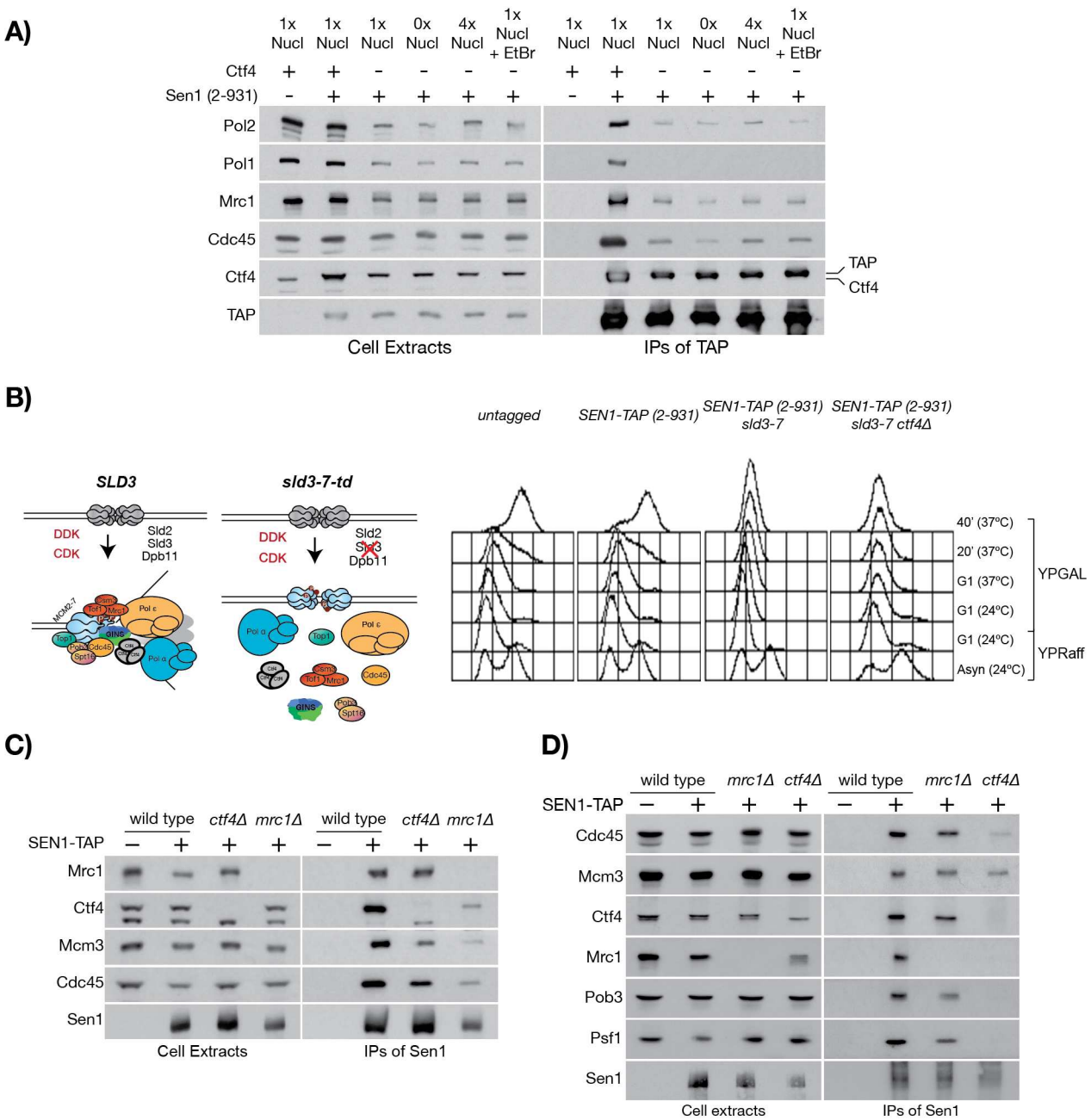

A)
1x Nucl
1x Nucl
1x Nucl
0x Nucl
4x Nucl
1x Nucl + EtBr
1x Nucl
1x Nucl
1x Nucl
0x Nucl
4x Nucl
1x Nucl + EtBr
Ctf4
Sen1 (2-931)
Pol2
Pol1
Mrc1
Cdc45
Ctf4
TAP
TAP
Ctf4
Cell Extracts
IPs of TAP
B)
untagged
SEN1-TAP (2-931)
SEN1-TAP (2-931) sld3-7
SEN1-TAP (2-931) sld3-7 ctf4Δ
SLD3
sld3-7-td
DDK
CDK
Sld2
Sld3
Dpb11
40' (37°C)
20' (37°C)
G1 (37°C)
G1 (24°C)
G1 (24°C)
Asyn (24°C)
YPGAL
YPRaff
C)
wild type
ctf4Δ
mrc1Δ
wild type
ctf4Δ
mrc1Δ
SEN1-TAP
Mrc1
Ctf4
Mcm3
Cdc45
Sen1
Cell Extracts
IPs of Sen1
D)
wild type
mrc1Δ
ctf4Δ
wild type
mrc1Δ
ctf4Δ
SEN1-TAP
Cdc45
Mcm3
Ctf4
Mrc1
Pob3
Psf1
Sen1
Cell extracts
IPs of Sen1

Figure S2. **Mrc1 and Ctf4 mediate Sen1 binding to the replisome** (related to Fig 2). **A)** Cells carrying the *GAL1-TAP-SEN1 (2- 931)* construct or the empty control were grown in YPGAL, arrested in G1 phase using α-factor and released in S phase for 20 min at 30°C. The samples were then used for IPs using TAP beads and treated with the indicated amount of nuclease or ethidium bromide (50 μg/ml). Ctf4 and TAP-Sen1 (2-931) have similar sizes and run closely in gel electrophoresis. **B)** (Left) Schematic representation of the system used in the experiment shown in Fig 2D; (Right) FACS profile of the experiment conducted. Cells were grown in YPRAF at 24°C, arrested in G1 and either harvested, or resuspended in YPGAL at 24°C for 35 min to induce the expression of Sen1 (2-931) and Ubr1, shifted to 37°C for 1 h to inactivate/degrade td-Sld3-7 and then released in S phase for 20 min at 37°C. **C)** Wild type, *ctf4*Δ and *mrc1*Δ cells, carrying a TAP-tagged or untagged allele of *SEN1*, were arrested in G1 and synchronously released in S phase for 30 min at 24˚C. Cell extracts were incubated with anti-TAP beads and the immunoprecipitated material was analysed by immunoblotting. **D)** Immunoblotting analysis of cell extracts and IP material from anti-TAP beads. The experiment was conducted an in **C)**, except cultures were incubated with formaldehyde before collection.

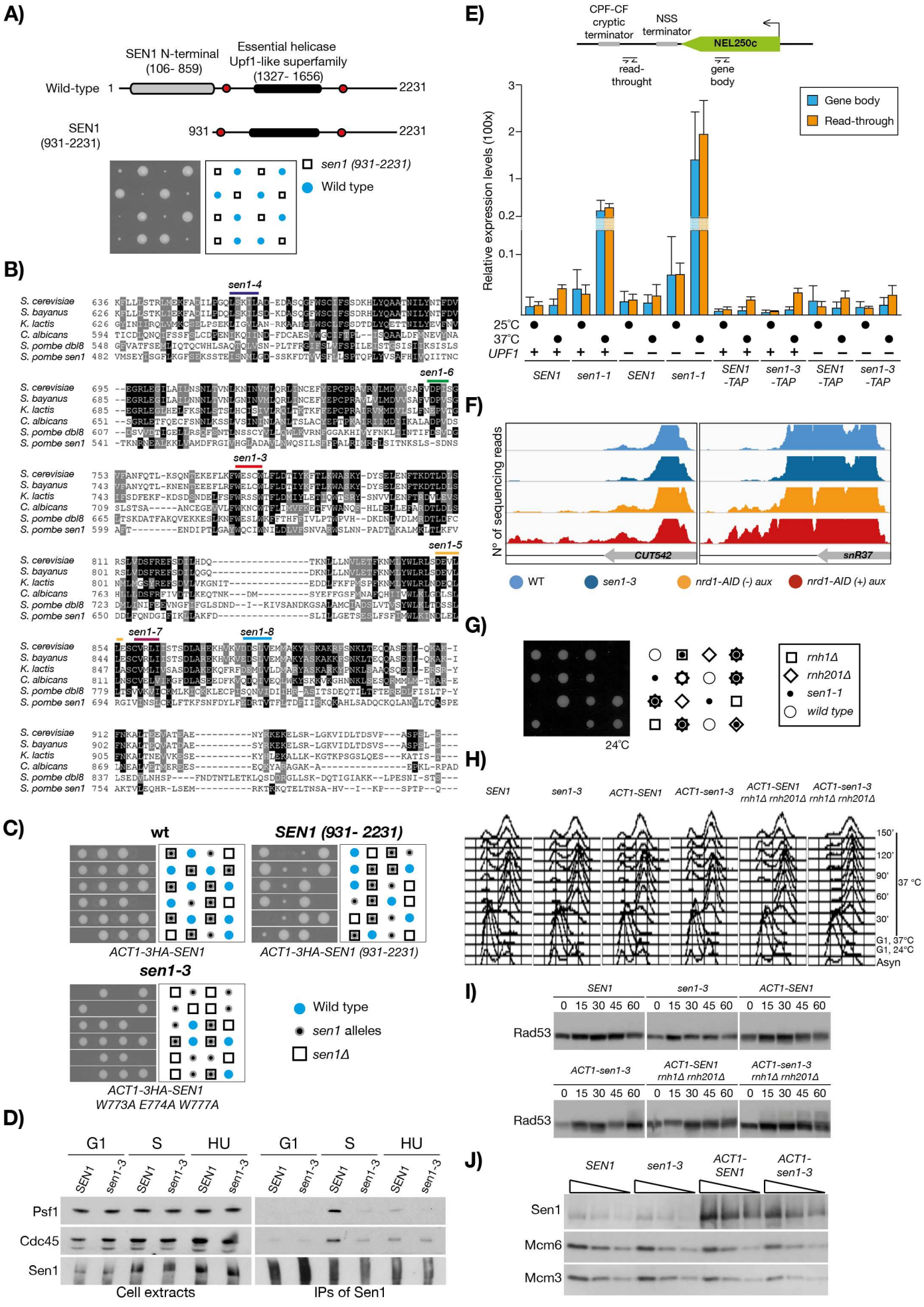

A)
SEN1 N-terminal (106- 859)
Essential helicase Upf1-like superfamily (1327- 1656)
Wild-type 1 2231
SEN1 (931-2231) 931 2231
sen1 (931-2231)
Wild type
B)
sen1-4
sen1-6
sen1-3
sen1-5
sen1-7
sen1-8
S. cerevisiae
S. bayanus
K. lactis
C. albicans
S. pombe dbl8
S. pombe sen1
C)
wt
SEN1 (931- 2231)
ACT1-3HA-SEN1
ACT1-3HA-SEN1 (931-2231)
sen1-3
ACT1-3HA-SEN1
W773A E774A W777A
Wild type
sen1 alleles
sen1Δ
D)
G1
S
HU
SEN1
sen1-3
Psf1
Cdc45
Sen1
Cell extracts
IPs of Sen1
E)
CPF-CF cryptic terminator
NSS terminator
NEL250c
read-throught
gene body
Relative expression levels (100x)
Gene body
Read-through
25°C
37°C
UPF1
SEN1
sen1-1
SEN1 -TAP
sen1-3 -TAP
F)
N° of sequencing reads
CUT542
snR37
WT
sen1-3
nrd1-AID (-) aux
nrd1-AID (+) aux
G)
24°C
rnh1Δ
rnh201Δ
sen1-1
wild type
H)
SEN1
sen1-3
ACT1-SEN1
ACT1-sen1-3
ACT1-SEN1 rnh1Δ rnh201Δ
ACT1-sen1-3 rnh1Δ rnh201Δ
150'
120'
90'
60'
30'
37 °C
G1, 37°C
G1, 24°C
Asyn
I)
Rad53
0 15 30 45 60
J)
Sen1
Mcm6
Mcm3

Figure S3. The N-terminal of Sen1 is important for cell growth and is conserved in yeasts (related to Fig 3 and 4)**. A)** Tetrad analysis of a diploid yeast strain carrying the *SEN1/sen1 (1-931Δ)* alleles. Plates were imaged after 5 days of growth on YPD at 24°C. **B)** Alignment of *Saccharomyces cerevisiae* Sen1 domain interacting with the replisome (636-931) with its orthologues from *Saccharomyces bayanus*, *Kluyveromyces lactis*, *Candida albicans* and *Schizosaccharomyces pombe dbl8* and *sen1*. The mutations used in the screen were selected to mutate conserved amino acids predicted to be on the surface (www.predictprotein.org). **C)** Analysis of the *ACT1-3HA-SEN1* mutants. Tetrad analyses were conducted on diploids yeast strains carrying *SEN1*/*sen1*Δ, and ectopically integrated *ACT1-3HA-SEN1* alleles at the *leu2-3,112* locus. Plates were imaged after 3 days of growth on YPDA at 24°C. **D)** Sen1-3 show greatly reduced interaction with RFs. Wild type and *sen1-3* cells were arrested in G1 and synchronously released for 30 min in fresh medium (S) or for 90 min in medium containing 0.2 M HU. At the indicated times, cultures were treated with formaldehyde before being collected. The cross-linked cell extracts and the immunoprecipitated material from anti-TAP beads were analysed by immunoblotting**. E)** (Top) Schematic representation of the gene analysed and the probes used to assess defects in transcription termination; (Bottom) RT-qPCR analysis of RNAs derived from the indicated strains. *NEL025c* is a non-coding region as described in (Wyers *et al.*, 2005). Cells were grown to exponential phase and incubated for 3 h at the indicated temperature before being collected. The signal is presented as the expression level relative to the housekeeping gene *ACT1* (triplicate biological repeats). **F)** Snapshots illustrating RNA Pol II density detected by CRAC on two NNS complex targets (one CUT and one snoRNA) in the indicated strains. An *nrd1-AID* strain grown in the presence or absence of auxin is included as a control for transcription termination (dataset from (Candelli *et al.*, 2018)). **G)** *sen1-1* is lethal in the absence of *RNH201* and *RNH1*. Examples are shown of tetrad analyses conducted from yeast diploids strains with the *SEN1*/*sen1-1 RNH1/rnh1*Δ and *RNH201/rnh201*Δ genotype. Plates were imaged after 4 days of growth on YPDA at 24°C. **H)** FACS analysis of the cell cycle progression in cells *SEN1*, *sen1-3, ACT1-SEN1, ACT1-sen1-3, ACT1-SEN1 rnh1*Δ *rnh201*Δ and *ACT1-sen1-3 rnh1*Δ *rnh201*Δ. Cells were grown to the exponential phase at 24°C, arrested in G1, shifted to 37°C for 1 hour in G1, and released in S phase at 37°C. The samples were collected at the indicated time points. **I)** *ACT1-sen1-3 rnh1*Δ *rnh201*Δ cells show activation of Rad53 during DNA replication at 37°C. Western blot analysis of samples taken from the experiment shown in **H). J)** Analysis of the protein levels of the alleles *SEN1*, *sen1-3*, *ACT1-SEN1* and *ACT1-sen1-3*, all carrying a 3HA tag, and loading controls Mcm3 and Mcm6.

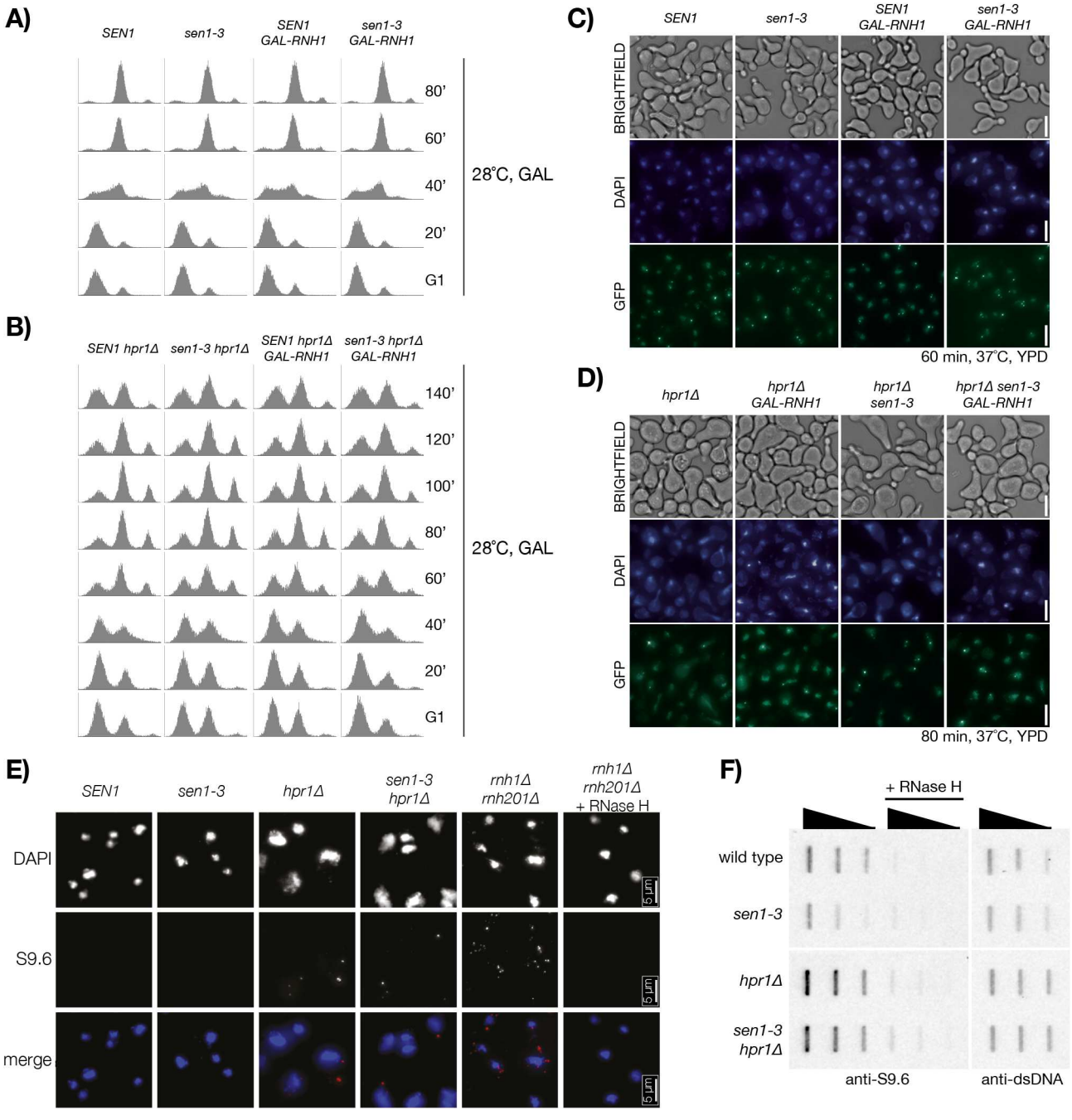

A)
SEN1
sen1-3
SEN1 GAL-RNH1
sen1-3 GAL-RNH1
80'
60'
40'
20'
G1
28°C, GAL
B)
SEN1 hpr1Δ
sen1-3 hpr1Δ
SEN1 hpr1Δ GAL-RNH1
sen1-3 hpr1Δ GAL-RNH1
140'
120'
100'
80'
60'
40'
20'
G1
28°C, GAL
C)
SEN1
sen1-3
SEN1 GAL-RNH1
sen1-3 GAL-RNH1
BRIGHTFIELD
DAPI
GFP
60 min, 37°C, YPD
D)
hpr1Δ
hpr1Δ GAL-RNH1
hpr1Δ sen1-3
hpr1Δ sen1-3 GAL-RNH1
BRIGHTFIELD
DAPI
GFP
80 min, 37°C, YPD
E)
SEN1
sen1-3
hpr1Δ
sen1-3 hpr1Δ
rnh1Δ rnh201Δ
rnh1Δ rnh201Δ + RNase H
DAPI
S9.6
merge
5 μm
F)
+ RNase H
wild type
sen1-3
hpr1Δ
sen1-3 hpr1Δ
anti-S9.6
anti-dsDNA

Figure S4. Analysis of the recombination and DNA:RNA hybrids in *hpr1*Δ and *hpr1*Δ *sen1-3* (related to Fig 4)**. A-D)** FACS analysis of the DNA replication dynamics and examples of the microscopy data of the experiments shown in Fig 4F are shown. Cells were grown to exponential phase in YPRAF at 28°C, arrested in G1, resuspended in YPGAL for 1 h and synchronously released in S phase in YPGAL (triplicate biological repeats). Scale bar = 5 μm **E)** Examples of the immunohistochemistry analysis of DNA:RNA shown in Fig 4G (triplicate biological repeats). Scale bar = 5 μm **F)** Analysis of R-loops *ex vivo***.** Cells were grown to exponential phase, arrested in G1 and then synchronously released in S phase for 30 min at 24°C. DNA:RNA hybrids double-stranded DNA were recovered in nuclease-free water and 1, 0.5 and 0.25 μg/μl dilutions of nucleic acid samples were prepared. The samples were then either treated with a commercially-sourced RNase H (or mock-treated), transferred onto nylon membrane and probed against using either the S9.6 antibody (that recognise R-loops) or an anti-dsDNA antibody.

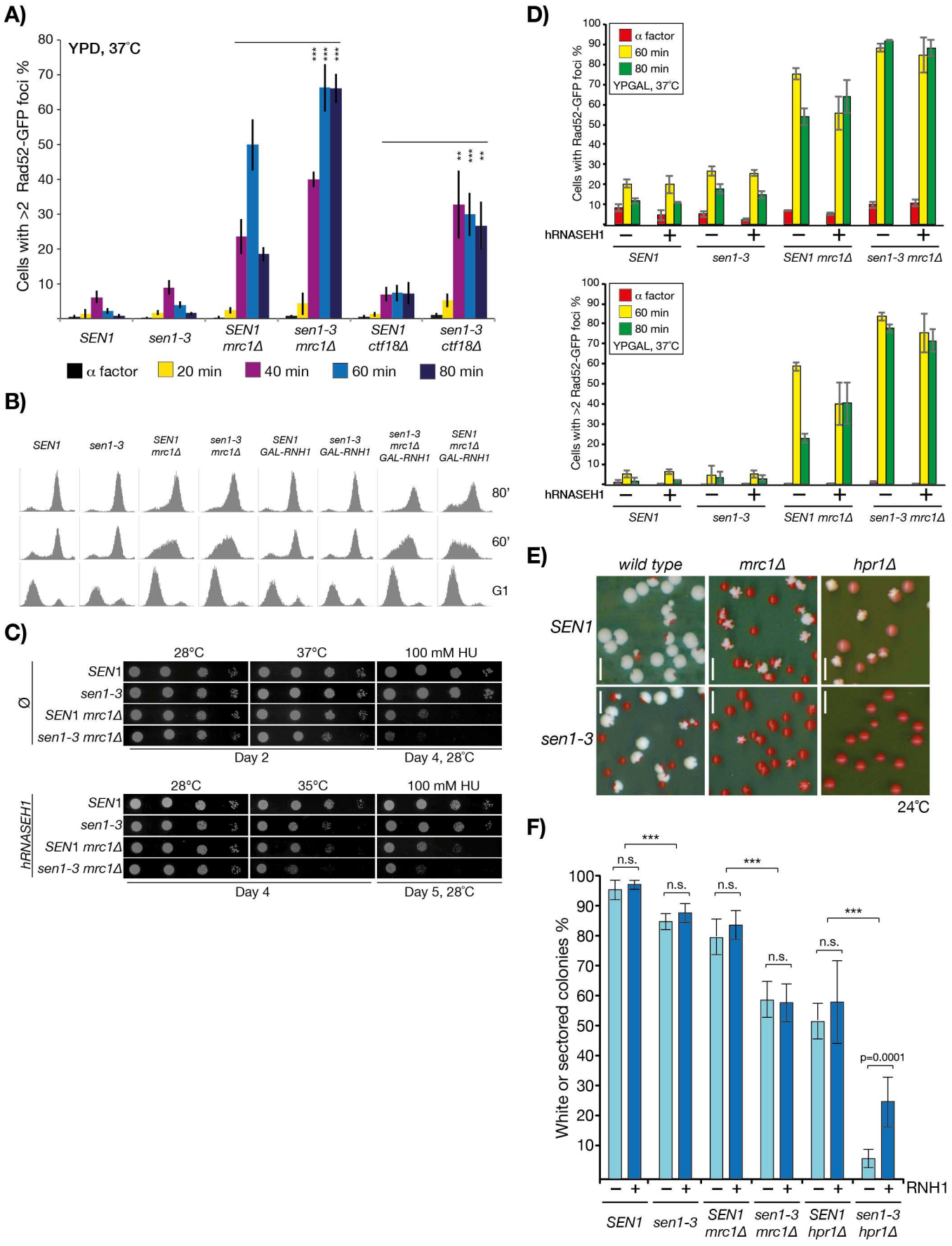

A)
YPD, 37°C
Cells with >2 Rad52-GFP foci %
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
SEN1 ctf18Δ
sen1-3 ctf18Δ
α factor
20 min
40 min
60 min
80 min
B)
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
SEN1 GAL-RNH1
sen1-3 GAL-RNH1
sen1-3 mrc1Δ GAL-RNH1
SEN1 mrc1Δ GAL-RNH1
80'
60'
G1
C)
28°C
37°C
100 mM HU
Ø
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
Day 2
Day 4, 28°C
28°C
35°C
100 mM HU
hRNASEH1
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
Day 4
Day 5, 28°C
D)
α factor
60 min
80 min
YPGAL, 37°C
Cells with Rad52-GFP foci %
hRNASEH1
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
E)
wild type
mrc1Δ
hpr1Δ
SEN1
sen1-3
24°C
F)
White or sectored colonies %
n.s.
p=0.0001
RNH1
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
SEN1 hpr1Δ
sen1-3 hpr1Δ

Figure S5. Overexpression of *RNH1* does not suppress the defects in *mrc1*Δ *sen1-3* (related to Fig 5)**. A)** Samples from experiments show in Fig 5D were scored for the presence of two or more foci per cell. ** p<0.05, *** p<0.01). **B)** FACS analysis of the experiment shown in Fig 5G. **C)** *hRNASEH1* overexpression does not suppress the defects observed in *sen1-3 mrc1*Δ cells. *hRNASEH1* was overexpressed from a 2 micron multicopy plasmid under the strong GPD^{TDH3} promoter. *SEN1*, *sen1-3*, *mrc1*Δ and *mrc1*Δ *sen1-3* cells were transformed with an empty or GPD^{TDH3}-*hRNASEH1* plasmid. Eight independent clones were pooled together and used for dilution spotting in medium lacking histidine, so to maintain the selective pressure for the plasmid. The cells carrying *GDP-hRNH1* grew more slowly and scans of their growth were taken at later times (Bottom panel). Serial dilution spotting (1:10) of the indicated strains is shown. **D)** *hRNASEH1* overexpression does not suppress the increase in recombination in *sen1-3 mrc1*Δ. Cell cultures were grown overnight at 24°C in medium lacking histidine to the exponential phase. Cells were diluted, resuspended in YPD, and left to grow for the length of one cell cycle. Cells were arrested in G1, shifted to 37°C for 1 h still with α-factor, and released in S phase at 37°C. Cells were taken at the indicated times, fixed, and analysed (triplicate biological repeats). **E)** Examples of the plasmid loss phenotype observed in the strains shown in Fig 5I (plasmid with 1 origin). Scale bar: 5 mm. **F)** The indicated strains, either carrying the *GAL1-RNH1* construct integrated at *leu2-3,112* or not, were transformed with the pRS315-*ADE2* plasmid. The experiment was performed as in Fig 5I, except that all media used contained galactose. n.s. = not significant, *** p<0.001.

II. Sen1 is a master regulator of transcription-driven conflicts.

In this manuscript I present the characterisation of the role of Sen1 in transcription-replication and transcription-transcription conflicts and the contribution of RNases H.

This work was a substantial part of my PhD and I have performed all the experiments and analysis presented in this study with the exception of the Sen1-AID RNAPII CRAC and its bioinformatic analysis presented in Figure 1. A detailed contribution of the other authors is included in the manuscript.

Sen1 is a master regulator of transcription-driven conflicts

Umberto Aiello[1], Drice Challal[1,2], Griselda Wentzinger[1], Armelle Lengronne[3], Rowin Appanah[4], Philippe Pasero[3], Benoit Palancade[1], Domenico Libri[1]*

* to whom correspondence should be addressed: domenico.libri@ijm.fr

[1] Université de Paris, CNRS, Institut Jacques Monod, F-75013 Paris, France
[2] present address : Institut de Biologie Intégrative de la Cellule UMR9198 CEA, CNRS, Université Paris-Saclay, F-91198 Gif-sur-Yvette, France
[3] Institut de Génétique Humaine, CNRS, Université de Montpellier, Montpellier, France
[4] Genome Damage and Stability Centre, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9RQ, UK

ABSTRACT

Cellular homeostasis requires the coordination of several machineries concurrently engaged on the DNA. Wide-spread transcription can interfere with other processes and transcription-replication conflicts (TRCs) threaten genome stability. The conserved Sen1 helicase terminates non-coding transcription, but also interacts with the replisome and reportedly resolves genotoxic R-loops. Sen1 prevents genomic instability but how this relates to its molecular functions remains unclear. We generated high-resolution, genome-wide maps of transcription-dependent conflicts and R-loops using a Sen1 mutant that has lost interaction with the replisome but is termination proficient. We show that Sen1 removes RNA polymerase II at TRCs within genes and the rDNA, but also at sites of transcription-transcription conflicts under physiological conditions, thus qualifying as a “master regulator of conflicts”. We demonstrate that genomic stability is only affected by Sen1 mutation when, in addition to its role at the replisome, termination of non-coding transcription or R-loop removal are additionally compromised.

Keywords: transcription; replication; transcription-replication conflicts (TRCs); Sen1; RNase H; R-loops; genome stability; non-coding transcription; H-CRAC

INTRODUCTION

The DNA is the shared workspace synchronously used by many cellular machineries that are essential for the correct expression, maintenance, repair, and transmission of the genetic information. Because they work concurrently, the orchestration of these activities must be accurately coordinated, both in time and space, to avoid interferences that might ultimately lead to mis-expression or corruption of the genetic content. Seemingly at odd with these necessities, transcription occupies the virtual integrity of the genome. RNA Polymerase II (RNAPII) transcribes largely beyond the limits dictated by apparent physiological significance, a phenomenon dubbed pervasive transcription. Robust and accurate mechanisms are required for limiting conflicts or solving them, but the actors involved, and their mode of action are not fully understood.

Transcription-replication conflicts (TRCs) are of marked interest in the crowded genomic landscape as they can generate genomic instability and jeopardize the faithful transmission of genetic information. The inherent stability of transcription elongation complexes is sufficient for inducing stalling of replication forks, a condition that has potential to generate DNA damage, in particular when associated to the formation of R-loops. These structures are characterized by a peculiar topological arrangement in which the nascent RNA associates to its DNA template, leaving unpaired the cognate DNA strand. R-loops have important physiological functions in the generation of antibody diversity, and other processes (for a review see: Feng et al., 2020), but their non-physiological accumulation is generally considered genotoxic.

The helicase Sen1 has a particular place in the orchestration of transcription and replication activities. Within the Nrd1-Nab3-Sen1 (NNS) complex, it has an essential role in controlling transcription termination at thousands of genes producing non-coding RNAs, some of which are stable and functional, such as snoRNAs involved in rRNA maturation, while others are unstable (Cryptic Unstable Transcripts, CUTs) and degraded by the nuclear exosome rapidly after transcription (Steinmetz et al., 2006; Hazelbaker et al., 2012; Porrua and Libri, 2013; Schaughency et al., 2014). Failures in NNS-dependent termination by depletion of Nrd1

has been shown to generate extended transcription events that affect the expression of neighboring genes, thus altering the overall transcriptional homeostasis of the cell (Schulz et al., 2013). Besides a role in limiting the chances of conflicts by restricting pervasive transcription, Sen1 has been proposed to work directly at sites of TRCs. Sen1 loss-of-function mutants, or strains in which the protein has been depleted, display genomic instability phenotypes revealed by increased mitotic recombination between direct repeats, synthetic lethality with DNA repair mutants and Rad52 *foci* accumulation (Mischo et al., 2011). These effects have been attributed to the defective resolution of R-loops at TRCs in the light of increased fork stalling at sites of convergent transcription and replication and the genomic co-localization of Sen1 and replication forks (Alzu et al., 2012). Indeed, increased R-loop levels have been detected in these Sen1 loss-of-function genetic backgrounds especially during S-phase (Mischo et al., 2011; San Martin-Alonso et al., 2021), which led to the proposal that Sen1, by virtue of its helicase activity, resolves R-loops that are formed at TRCs. However, in these mutant contexts, transcription termination of many ncRNA genes is affected, with potential effects on the phenotypes observed. Thus, in the absence of Sen1 it is conceivable that rather than (or in addition to) the defective resolution of constitutively formed R-loops, more R-loops are formed as a consequence of the generally higher transcriptional readthrough, which entails increased chance of conflicts. Although genomic instability was not observed in other mutants of the NNS complex that have termination defects (Costantino and Koshland, 2018; Mischo et al., 2011), it remains possible that the phenotypes associated to Sen1 loss-of-function originate from the synthetic association of increased transcriptional challenges and failure to resolve conflicts. Disentangling the contributions of these potential synthetic effects is paramount for understanding the function of Sen1 in maintaining genomic stability.

We have recently reported the physical interaction of Sen1 with the Ctf4 and Mrc1 replisome components and characterized a mutant, *sen1-3*, that loses this interaction (Appanah et al., 2020). *Sen1-3* cells have a minor growth phenotype and, importantly, no transcription termination defects at NNS target genes. However, this mutation induces lethality

in the absence of the two yeast RNases H, Rnh1 and Rnh201/202/203, which, together with other genetic interactions with mutants of the fork stalling signaling pathway (Appanah et al., 2020), underscores the physiological relevance of the interaction of Sen1 with the replisome. Importantly, RNase H1 and H2 are redundantly involved in the degradation of the RNA moiety of R-loops, which might mechanistically underlie a functional connection with Sen1 at TRCs.

In the course of a separate study we have shown that Sen1 is required for a back-up mechanism of RNAPIII release when primary termination has failed (Xie et al., 2021). The interactions of Sen1 with the replisome and RNAPIII are mutually exclusive, which implicates the existence of two distinct Sen1-containing complexes presumably with different functional roles. Interestingly, however, both interactions are impaired by the *sen1-3* mutation, which likely alters a shared region of interaction (Xie et al., 2021). These findings indicate that the *sen1-3* mutant allows untangling the function of Sen1 in NNS-dependent termination from its functions at the replisome and RNAPIII transcription.

Here we first addressed the functional impact of the *sen1-3* mutation on transcription-replication conflicts under conditions in which neither transcription nor replication are altered. Prompted by the strong genetic interaction with RNases H we also studied the role of these enzymes at sites of conflicts and the impact of R-loops. To this aim we generated high resolution transcription maps in different phases of the cell cycle in *sen1-3* cells and in the absence of RNases H. We also devised a novel methodology to detect R-loops *in vivo* with high sensitivity and unprecedented resolution. We show that Sen1 is required for the efficient removal of RNAPII at many TRC sites within genes, but also at the rDNA, where it collaborates with RNases H. Surprisingly, when non-coding transcription is correctly terminated, loss of the interaction of Sen1 with the replisome does not cause increased R-loop accumulation, mitotic recombination or DNA damage as observed in *sen1* loss-of-function mutants. We demonstrate that increased DNA damage observed in these mutants requires both the lack of Sen1 interaction with the replisome and the transcription termination defects at non-coding RNA genes. Interestingly, we show that Sen1 also functions at many other genomic sites to remove RNAPII at sites of conflicts with RNAPIII and possibly RNAPI. We propose a model according

to which Sen1 functions as a “master regulator of conflicts” in the genome, and we demonstrate that this function is independent from its role in terminating non-coding transcription.

RESULTS

Sen1 depletion leads to major genome-wide alterations in the transcriptome

To assess directly the possible contribution of transcription termination defects to the genomic instability phenotypes observed in Sen1 loss-of-function mutants, we first directly gauged the extent of alterations in coding and non-coding RNAPII transcriptional activity under defective Sen1 function, which was previously investigated only to a limited extent (Schaughency et al., 2014). In this former report Sen1 was depleted by the anchor away methodology (Haruki et al., 2008), which, in our hands, did not induce major growth defects, possibly because of partial Sen1 depletion. For consistency with the data in this and other studies (Costantino and Koshland, 2018; San Martin-Alonso et al., 2021), we generated high resolution transcription maps using RNAPII CRAC (Crosslinking Analysis of cDNAs, Bohnsack et al., 2012; Candelli et al., 2018) upon depletion of Sen1 by the auxin degron system (Nishimura et al., 2009). In the CRAC methodology, the nascent RNA is crosslinked to Rpb1, the largest subunit of RNAPII, which is purified under denaturing conditions, thus limiting co-purification of associated, non-nascent transcripts. Sequencing of the crosslinked RNA provides the position of the polymerase with directionality and high resolution.

Depletion of Sen1 by addition of auxin induced the expected transcription termination defects at canonical NNS targets (CUT and snoRNA genes, Figures S1A, left panel, and S1B), but not at mRNA coding genes (Figure S1A, right panel), consistent with the Schaugency et al. study (Schaughency et al., 2014).

Defective RNAPII release at NNS targets could ultimately result in a global redistribution of polymerases with increased persistency in some regions, as well as depletions in others, possibly because of transcription interference. To estimate the occurrence of global transcriptional changes in an unbiased manner, we first computed the RNAPII CRAC signal in intergenic regions divided in 200 nt, non-overlapping windows with the exclusion of mRNA-coding genes, the rDNA and tRNA genes to first focus on regions of direct Sen1 action. Scatter plots of these values revealed a dramatic alteration of intergenic transcription upon Sen1 depletion compared to wild-type cells (Figures 1A, compare left and right). Similar

transcription alterations were observed upon auxin-dependent depletion of another NNS component, Nrd1, indicating that they are linked to a transcription termination defect (Figure S1C).

Figure 1

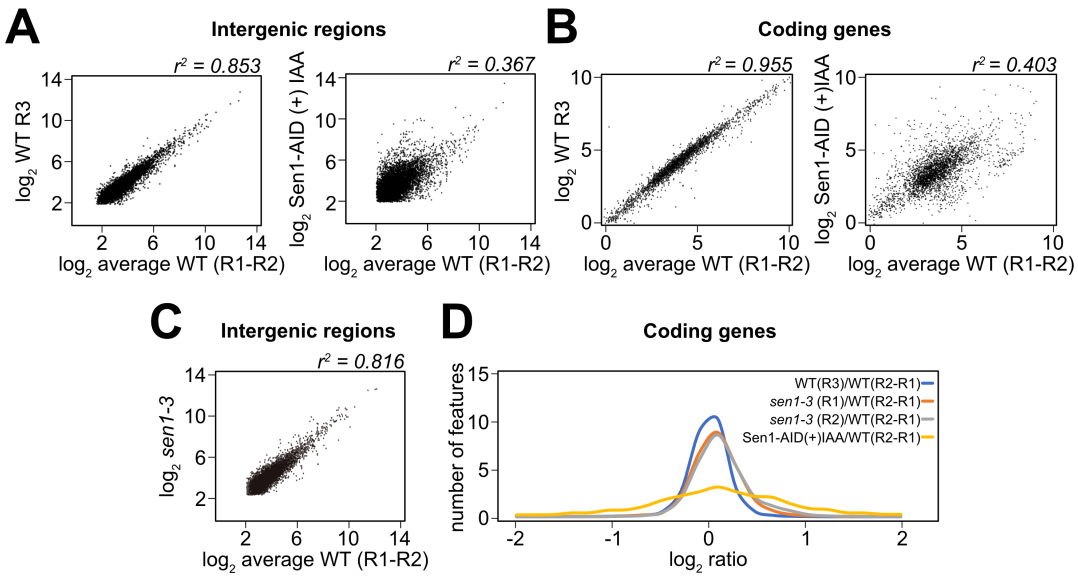


Figure 1: Major alterations in the transcriptome upon Sen1 depletion

A) Scatter plots of RNAPII CRAC log_2 values computed in non-overlapping 200 nt bins relative to the average of two wild-type replicates (WT, R1-R2) used as a common reference in all panels. Only the W strand was used to exclude background signals derived from contaminating rRNA. mRNA-coding and tRNA genes have also been excluded from this analysis. The scatter plot of RNAPII CRAC signals derived from Sen1 depleted cells is presented (right), together with the scatter plot of another wild-type replicate (left, WT R3) for comparison and visual assessment of reproducibility. Bins containing a low number of reads (log_2<2) have been excluded from the analysis for clarity. **B)** As in A but RNAPII CRAC average signal (log_2) from mRNA coding genes have been computed. **C)** As in A, right panel, intergenic RNAPII CRAC values from *sen1-3* cells are compared to the wild-type reference. **D)** Distribution of log_2 ratios of average RNAPII CRAC signals detected on mRNA coding genes. Data from two independent *sen1-3* replicates, evaluated relative to the wild-type reference. An additional wild-type replicate is shown for comparison together with the distribution obtained upon Sen1 depletion.

Deregulation of non-coding transcription might induce significant effects on the protein-coding transcriptome, which are susceptible to influence normal cellular physiology. Therefore, we also analyzed the transcription levels of mRNA-coding genes in Sen1-AID strains by RNAPII CRAC. The gene expression program was also significantly modified upon Sen1 depletion (Figure 1B). These effects were generally recapitulated upon depletion of Nrd1, the two datasets showing a high level of correlation (r^2=0.726, Figure S1D). Because we did not observe termination defects at genes with altered expression (and at mRNA genes in general, Figure S1B, first panel), it is likely that these effects are at least partially due to a cellular response to the perturbations introduced by Sen1 (and Nrd1) depletion. Notably, many stress genes are activated and ribosomal protein genes downregulated, possibly suggesting the occurrence of a marked stress response (see Discussion).

These results demonstrate that loss of full Sen1 function has a major impact on the distribution of transcription events genome-wide and on gene expression. These changes are expected to alter the wild-type landscape of TRCs and might contribute significantly to the genomic instability phenotypes observed in Sen1 loss-of-function mutants.

Interaction of Sen1 with the replisome is required to solve TRCs in the 5'-end of genes.

In the light of the above considerations, we sought to analyze the role of Sen1 in genomic stability in a context that would be more physiological in terms of transcriptional landscape. Genetic data strongly support the notion that the role of Sen1 at TRCs is mechanistically linked to its interaction with the replisome, and therefore we turned to the use of the *sen1-3* mutant that loses this interaction, but is proficient for NNS termination (Appanah et al., 2020). In sharp contrast to what observed upon depletion of Sen1, affecting the interaction of Sen1 with the replisome does not alters significantly the coding and non-coding transcriptome (Figures 1C and 1D), validating the use of the *sen1-3* mutant for specifically focusing on the role of Sen1 at TRCs.

Figure 2

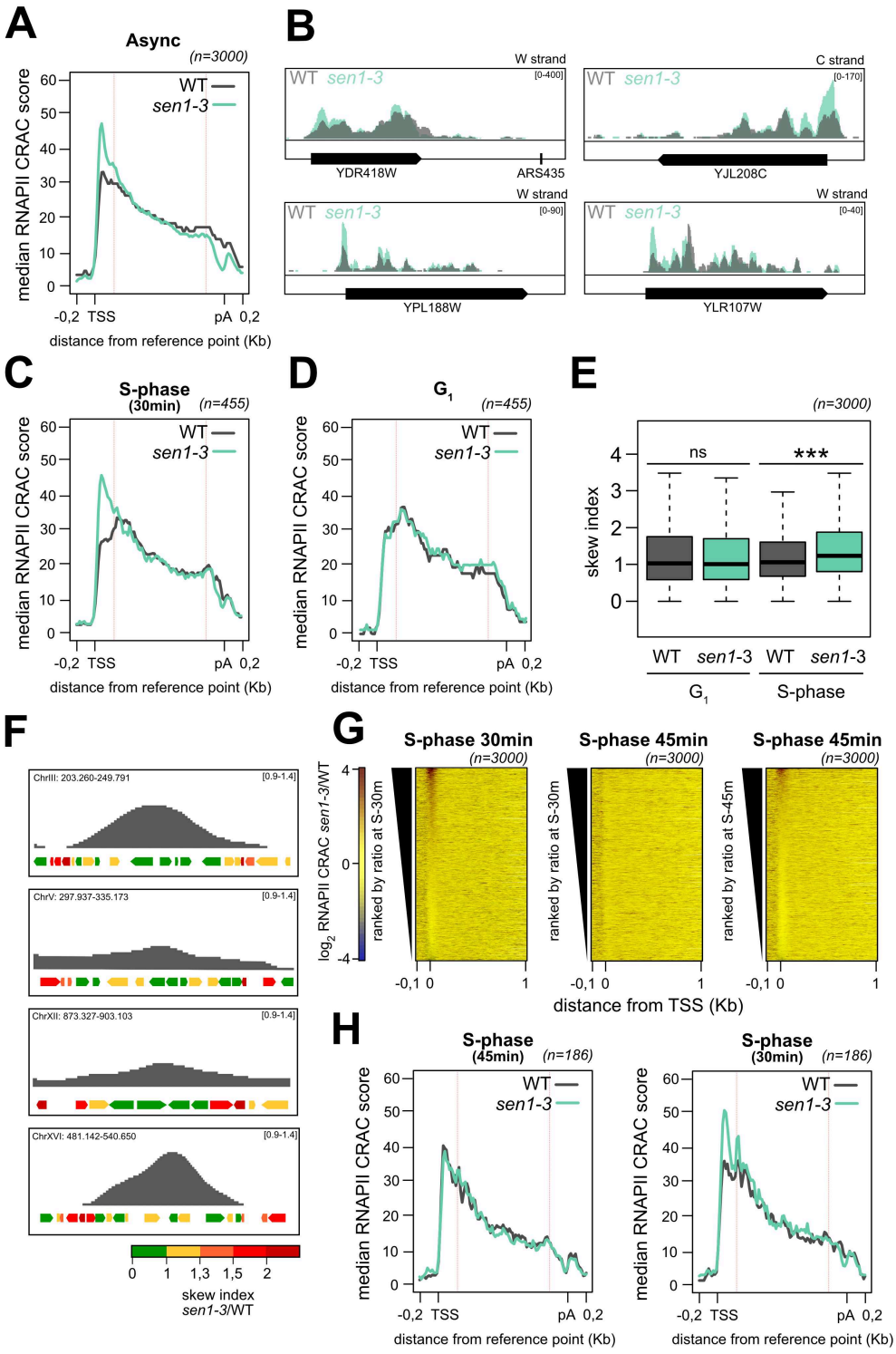

A
Async
(n=3000)
WT
sen1-3
median RNAPII CRAC score
-0,2 TSS
pA 0,2
distance from reference point (Kb)
B
W strand
C strand
WT sen1-3
[0-400]
[0-170]
YDR418W
ARS435
YJL208C
[0-90]
[0-40]
YPL188W
YLR107W
C
S-phase
(30min)
(n=455)
D
G1
(n=455)
E
(n=3000)
ns
***
skew index
WT
sen1-3
G1
S-phase
F
ChrIII: 203.260-249.791
ChrV: 297.937-335.173
ChrXII: 873.327-903.103
ChrXVI: 481.142-540.650
[0.9-1.4]
0 1 1,3 1,5 2
skew index
sen1-3/WT
G
S-phase 30min
S-phase 45min
S-phase 45min
(n=3000)
log2 RNAPII CRAC sen1-3/WT
ranked by ratio at S-30m
ranked by ratio at S-45m
distance from TSS (Kb)
H
S-phase
(45min)
(n=186)
(30min)

Figure 2: Sen1 promotes release of RNAPII from the 5'-end of genes undergoing replication.

A) Metagene analysis of the RNAPII distribution at coding genes aligned at their Transcription Start Site (TSS) and at their poly-Adenylation (pA) site in wild-type (WT) and *sen1-3* cells grown asynchronously. The 3000 genes with the highest expression (as determined by RNAPII CRAC) have been included in the analysis. Values on the y-axis correspond to the median coverage. Genes are only scaled in the interval delimited by red lines. **B)** Integrative Genomics Viewer (IGV) screenshots of representative examples of coding genes illustrating RNAPII accumulation in *sen1-3* cells. The overlap of RNAPII read coverage in wild-type (WT, grey) and *sen1-3* cells (aqua green) is shown. The scale and the strand (W for Watson, C for Crick) are indicated in brackets. **C)** Metagene analysis as in A, but for the genes with the highest *sen1-3*/WT skew index ratio (i.e. higher than the mean plus one standard deviation). Analysis performed in cells synchronously released in S-phase and collected 30 min after replication onset. **D)** Metagene analysis performed on the same gene set as in C but for cells arrested in G_1 by α-factor. **E)** Comparison of the gene skew ratio for the indicated strains in G_1-arrested cells or in cells synchronously released in S-phase and collected 30 min after replication onset. Analysis preformed on the 3000 most expressed genes as in A. *** p<0.001**. F)** Integrative Genomics Viewer (IGV) representative screenshots of replicons as detected by DNA copy number analyses for cells synchronously released in S-phase and collected after 30 min. Genes in each region were coloured according to their skew index *sen1-3*/WT ratio as indicated at the bottom of the panel. Genes with the highest skew index ratio (red and orange) are located preferentially at the borders of each replicon, where replication is most likely to be ongoing. **G)** Heatmap analyses representing the $\log_2$ fold change (FC) of the RNAPII CRAC signal in the *sen1-3* mutant relative to the WT for mRNA coding genes aligned at their Transcription Start Site (TSS) in S-phase at 30 min and 45 min as indicated. Genes were ranked, as indicated, according to the signals detected in the first 200 nt after the TSS at the 30 min (left and central panel) or 45 min time point (right panel). **H)** Left: Metagene analysis as in Figure 2C, but on cells collected 45 min after replication onset and on a group of genes already replicated at the indicated time point. Right: the same group of genes are undergoing replication at 30 min after release in S-phase and show the characteristic 5'-end skew in *sen1-3* cells.

Failure to resolve or to avoid a TRC is expected to result in slowing down or stalling of a replication fork but also to induce the accumulation of RNAPII at the site of conflict. We focused on the transcription side of the conflict and reasoned that if Sen1 is recruited at the replisome to terminate conflicting transcription, in its absence RNAPII should accumulate at these sites, thus providing a signature of Sen1-dependent conflicts.

Assuming that conflicts depend, at least to some extent, on the stochastic encounters of the two machineries, they can be expected to be more frequent in genomic regions with

inherently higher RNAPII occupancy. In this perspective, a marked asymmetry of RNAPII distribution is clearly observed for yeast genes in wild-type cells, with higher levels in the 5'-end (Figure S1A, right panel; see also Churchman and Weissman, 2011; Mayer et al., 2011) where the elongation complex is known to pause. Note that the general 5'-end proximal pausing is an inherent feature of the transcription process that is independent of replication as it is observed also in cells arrested in G_1 (see below, Figure 2D).

We therefore first analysed the distribution of RNAPII on yeast genes with a particular focus on the regions of transcriptional pausing. An alteration of the RNAPII profile was clearly observed by metasite analyses and by the inspection of individual genes in asynchronously growing *sen1-3* cells, with an increase in occupancy in the 5'-end that gradually decreases in the 3'-end (Figures 2A and 2B). In some cases, increased occupancy was also observed at other sites of RNAPII pausing (e.g., see *YDR418W* in Figure 2B). This pattern is not compatible with increased transcription initiation at a set of genes, which would result in a homogeneous, increase of the RNAPII CRAC signal all along these genes. Rather, it points to the occurrence of increased RNAPII pausing, mainly in the 5'-end of genes, possibly due to defective RNAPII release at sites of conflicts when Sen1 is absent from the replisome.

To assess the dependency on ongoing replication, RNAPII CRAC was performed using cells arrested in G_1 and synchronously released in S-phase. The progression of replication was analysed in the same cells by DNA copy number analyses. At 30 minutes after the release in S-phase, abnormal RNAPII occupancy was observed in *sen1-3* cells (Figure S2A), recapitulating what observed in asynchronous cells and, consistent with the hypothesis of a replication-dependent phenomenon, only a subset of genes was affected (see below, Figure 2G). For a quantitative analysis of these observations, we calculated a skew index, defined as the ratio of the RNAPII CRAC score in the 5'-end [TSS; TSS+200] to the signal in a downstream region of identical length [TSS+300; TSS+500]. The skew index is expected to be poorly sensitive to changes in transcription levels, which might slightly vary in the different strains. For more robust analyses, we focused on the subset of most affected genes, identified by computing the ratio of the skew indexes from the *sen1-3* and the wild-type cells and

selecting features with ratios one standard deviation over the mean. This resulted in a set of 455 genes, with a marked RNAPII CRAC signal increase in *sen1-3* cells slightly upstream of the canonical 5' peak detected in the wild-type (Figure 2C). Most importantly, increased RNAPII 5' occupancy was not observed in the absence of replication, when cells were arrested in G_1 by the addition of alpha factor, as demonstrated by the pattern of RNAPII distribution (compare Figures 2C and 2D) and the distributions of skew indexes, which were significantly different only in S phase (Figure 2E). This conclusion also holds genome-wide, when considering the set of most expressed 3000 genes (Figure S2B). These findings indicate that replication is required for the increased occupancy of RNAPII observed in the 5'-end of genes in *sen1-3* cells, supporting the notion that RNAPII is not released efficiently at TRCs when Sen1 cannot interact with the replisome.

If the replication-dependent RNAPII increase in the 5'-end of genes is linked to TRCs, the group of affected genes should be located in regions where replication is ongoing. To address this point, we divided all genes in 5 different groups based on the skew index ratio and monitored their distribution along the chromosomes in relation to the position of the replicative forks as detected by DNA copy number analyses in the same experiment. Because of population heterogeneity, the distribution of normalized DNA levels (Figures 2F and S2C, see Materials and Methods) is linked to the probability that a given sequence has already been replicated (close to the center of the replicon) or is undergoing replication (close to the periphery) at the analyzed time point. The genes with the highest skew index ratios were flanking or included in virtually every active replicon but were not distributed randomly. As shown in Figures 2F and S2C, we found that they are clearly preferentially positioned towards the borders, i.e., in regions where replication is most likely to be ongoing.

Genes in both a co-directional (CD) and head-on (HO) orientation relative to the direction of replication were equally found to be affected (Figures 2F, S2C and S2D), as also shown by measuring the distance of the closest origins generating HO or CD replication for each affected gene (Figure S2E). This suggests that Sen1 can solve both kinds of conflicts by binding to the replisome.

The position of TRCs is expected to change as replication progresses. Therefore, we generated additional RNAPII CRAC transcription maps at 45 min after release in S-phase and compared the results to the first transcription map that was generated at an earlier, 30 min time point. Analysis of the RNAPII CRAC signal generated at a later replication time point recapitulated the phenotype observed at the 30 min time point (Figure S2F and 2G, right panel), yielding a group of 439 affected genes (Figure S2G) selected following the aforementioned criteria. However, when the genes ranked for increased RNAPII occupancy at the 30 min time point were monitored for RNAPII CRAC signals at the 45 min time point (Figure 2G, compare left and middle panel), a very poor overlap, if any, was observed, consistent with the notion that TRCs are restricted to the set of genes undergoing replication at a given time point.

If RNAPII 5'-end accumulation is due to TRCs, genes that have already undergone replication at a given time point should not display increased RNAPII 5' persistency. Consistent with this notion, RNAPII CRAC analysis at a set of 186 genes selected for having been replicated at the 45 min time point according to our replication maps revealed no increase in RNAPII persistency at the 5'-end of genes in *sen1-3* cells (Figure 2H, left panel).

Increased persistency of RNAPII at many genomic sites might affect replication. In agreement with the presence of a challenged replication environment, the size of early replicons as assessed by DNA copy number analysis was found to be significantly smaller in *sen1-3* cells relative to the wild-type (Figure S2H, left panel). Interestingly, the same analysis also revealed the premature activation of late origins in the *sen1-3* mutant (Figure S2H, right panel), which compensates the general diminished fork progression, thus explaining the similar length of S-phase observed in *sen1-3* and wild-type cells (Appanah et al., 2020).

Together, these findings indicate that ongoing replication is required for the accumulation of RNAPII in the 5'-end of a set of genes, which we propose to be due to the failure to remove RNAPII from sites of TRCs when Sen1 cannot interact with the replisome. They also suggest that TRCs are frequently occurring even under physiological conditions and that the failure to efficiently prevent or resolve them alters the replication program.

Association of Sen1 with the replisome is required for limiting RNAPII accumulation at the ribosomal replication fork barrier in S-phase.

The ribosomal DNA Replication Fork Barrier (rRFB) is a site where one replication fork stalls upstream of the DNA-bound Fob1 protein at the 3'-end of ribosomal DNA repeats (Kobayashi, 2003). This ensures that each rDNA repeat is being replicated in a co-directional fashion with RNAPI and RNAPIII transcription (Figure 3A).

Figure 3

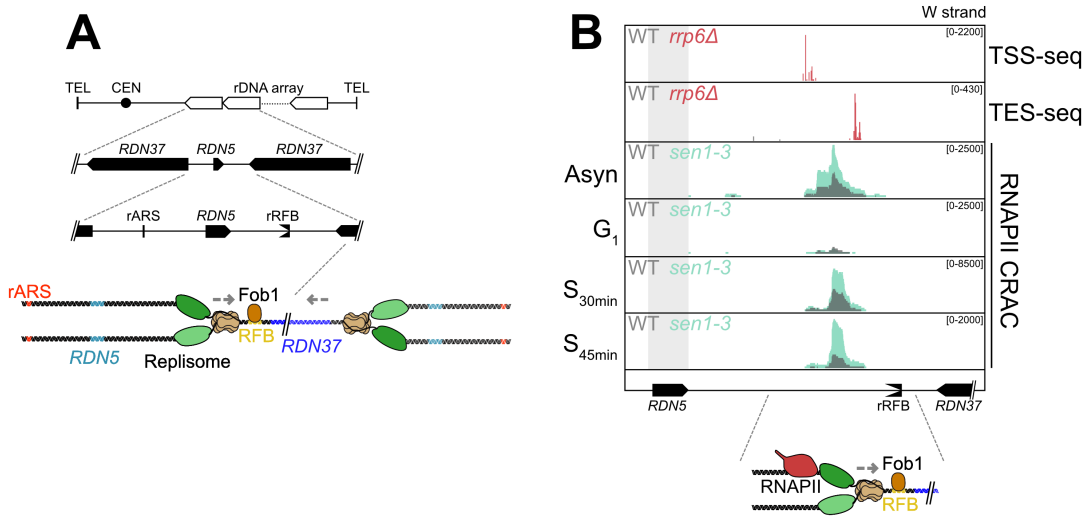


Figure 3: Association of Sen1 with the replisome is required for limiting TRCs at the replication fork barrier

A) Schematic representation of the budding yeast rDNA *locus* on ChrXII. The position of the replication origin (rARS) and of the rRFB are indicated relative to the *RDN37* and *RDN5* genes. A cartoon illustrates the process of replication in this region and the function of the Fob1 protein at the rRFB. The direction of progression for each replisome is indicated as a grey dashed arrow. **B)** Screenshots illustrating the distribution of the RNAPII CRAC signal density around the rRFB. Every track contains overlapped signals to compare either RNAPII CRAC levels in WT (grey) and *sen1-3* cells (aqua green) or TES-seq and TSS-seq signals for WT (grey) and *rrp6*Δ (red) as indicated. The TSS and the TES signals of the RNA produced by the fork-trailing transcription (first two tracks) are only visible for *rrp6*Δ cells because the RNA produced is degraded by the exosome and poorly detected in WT cells. Asyn: asynchronously growing cells; G_1: cells arrested in G_1; S_{30min}, S_{45min}: cells synchronously released in S-phase for 30 min and 45 min respectively. At the bottom of the panel, a cartoon illustrates the likely relative position of RNAPII and the fork.

Monitoring RNAPII by CRAC in the rRFB region in asynchronous cells revealed the existence of a non-annotated transcription unit located upstream and in close proximity of the rRFB, which generates a cryptic unstable transcript that can only be detected in an exosome-defective, *rrp6*Δ background as revealed by mapping its 5'- and 3'-ends (TSS and TES, respectively) (Figure 3B). Transcription in this region might generate a co-directional conflict with forks stalled at the rRFB and be of interest for our analysis.

Interestingly, the RNAPII CRAC signal was close to background in G_1-arrested cells yet was markedly visible both at the early (30 min release in S-phase) and late (45 min release) replication time points (Figure 3B), indicating that accumulation of RNAPII only occurs during or after replication, possibly as a consequence of fork passage or stalling.

Most importantly, the RNAPII CRAC signal was found to be considerably increased in *sen1-3* cells indicating that the interaction of Sen1 with the replisome is required for releasing RNAPII at this site while in close proximity with a replication fork. Increased RNAPII signal in *sen1-3* cells was not due to increased rDNA copy number in this strain as verified by qPCR (data not shown). Analysis of published ChIP-exo data (Rossi et al., 2021) confirmed the specific presence of Sen1 at the DNA in close correspondence with the RNAPII CRAC peak (Figure S4A). The RNAPII peak is located roughly 100 nt from the edge of the rRFB, which is hardly compatible with Fob1 roadblocking RNAPII as we have previously shown that the elongation complex stalls much closer (-10 to -15 nt) to DNA bound factors (Candelli et al., 2018; Colin et al., 2014). Rather, the position of the RNAPII peak is compatible with one co-directional replication fork derived from the closest ARS filling the gap between the RNA polymerase and the Fob1-bound rRFB (see scheme in Figure 3B). We cannot exclude that a head-on conflict occurs with the fork originating from the distal ARS and progressing in the opposite direction relative to transcription (Figure 3A), although this would imply distal fork stalling for undetermined reasons after crossing the rRFB.

Together these data provide evidence for replication-dependent RNAPII accumulation, most likely in the wake of the replication fork stalled at the rRFB. Importantly, they also demonstrate

that the interaction of Sen1 with the replisome is required for the efficient release of RNAPIIs paused in close proximity with the replisome.

R-loops are detected at the rRFB by H-CRAC

Several studies have linked mutation or depletion of Sen1 to the accumulation of R-loops at sites of conflicts, which has been proposed to generate genomic instability. However, as highlighted above, accumulation of R-loops in the absence of general Sen1 functions might depend on or be exacerbated by major alterations in the transcriptional load that challenges replication. The requirement of RNase H activity for the viability of *sen1-3* cells suggests that degradation of R-loops is essential in at least some genomic locations when Sen1 cannot interact with the replisome. Therefore, we decided to generate genome-wide maps of R-loops taking advantage of the unprecedented benefits offered by the *sen1-3* mutant.

Currently available tools to produce genome-wide R-loops maps have limited resolution and often lack directionally. These techniques rely on immunoprecipitation of DNA:RNA hybrids by the S9.6 antibody after nucleic acid extraction (DRIP-seq and related techniques), or the monitoring of catalytically-dead RNase H1 occupancy by chromatin immunoprecipitation (R-ChIP) (for a review see Chédin et al., 2021). These approaches do not always provide consistent outputs (Chédin et al., 2021) and we also feared that their resolution would not be sufficient for an integration with our RNAPII CRAC data. Therefore, we devised an alternative method: we reasoned that since RNase H binds and degrades the RNA moiety of DNA:RNA heteroduplexes, it should be possible to catch it in action at its targets *in vivo* by UV-crosslinking. Purification of the enzyme under denaturing conditions as in CRAC should allow sequencing of the crosslinked RNA for a sensitive and high-resolution detection of the hybrids. A similar strategy was successfully used to detect exosome targets (Delan-Forino et al., 2017), and should provide *in vivo* data obviating alterations in R-loop metabolism that might occur by expressing RNase H1 catalytic mutants. H-CRAC (for RNase H - CRAC) experiments performed with both RNase H1 and RNase H2 provided very reproducible and similar outputs, despite revealing some specificities (Figure 4A and S3A, see

below). This was expected in the light of the known redundancy of these enzymes in R-loop degradation.

Figure 4

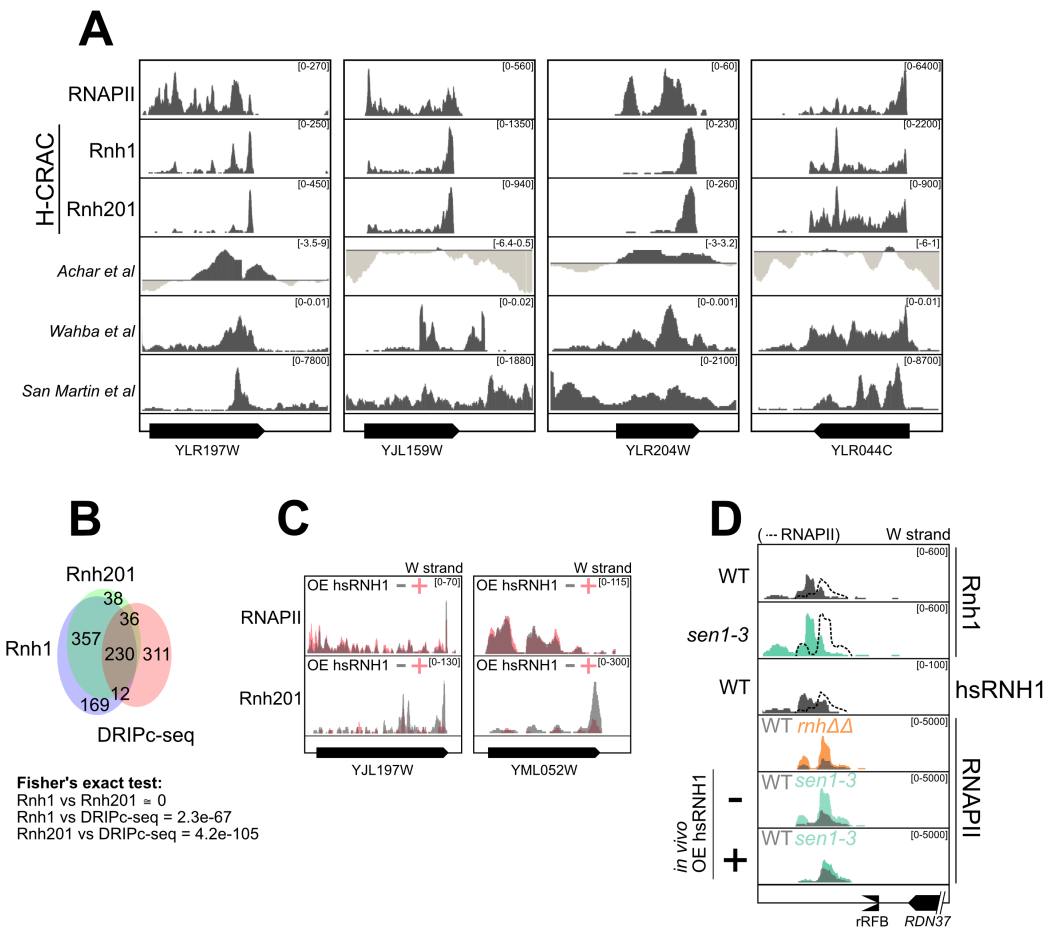


Figure 4: RNase H partakes in Sen1-dependent RNAPII release at the RFB

A) Snapshots of individual genes illustrating the comparison between DRIP datasets (Achar et al., 2020; San Martin-Alonso et al., 2021; Wahba et al., 2016) and H-CRAC signals obtained from Rnh1 and Rnh201. RNAPII CRAC is also shown for evaluating the R-loop signals relative to transcription. For the directional DRIP-seq (San Martin-Alonso et al., 2021), the H-CRAC and the RNAPII CRAC only the strand of the target gene is shown. **B)** Overlap of the genes containing the highest signals (levels higher than the mean plus one standard deviation) defined by the directional DRIP-seq (San Martin-Alonso et al., 2021) and by H-CRAC (Rnh1 or Rnh201). The significance of the overlap between H-CRAC and the DRIP-seq dataset was calculated with Fisher's exact tests and is indicated at the bottom.

C) Representative examples of decrease in H-CRAC (Rnh201) signals upon *in vivo* overexpression of hsRNH1. RNAPII occupancy by CRAC in both conditions is also shown to underscore that reduction of H-CRAC signal is not due to decreased transcription. **D)** H-CRAC and RNAPII CRAC at the rRFB in different genetic backgrounds as indicated: R-loops accumulation at the rRFB as detected by Rnh1 and hsRNH1 H-CRAC. The position where RNAPII accumulates in the corresponding background is indicated by a dashed curve. Increased RNAPII occupancy is observed at the rRFB by RNAPII CRAC in *rnh1Δ rnh201Δ* (*rnhΔΔ*) cells. RNAPII increased occupancy is suppressed by *in vivo* overexpression of hsRNH1.

To assess the reliability of our approach, we gauged the validity of the proposed landmarks for R-loop detection (Chédin et al., 2021). We first ectopically expressed in yeast a sequence derived from the mouse AIRN gene that was demonstrated to form R-loops *in vitro* and *in vivo* (Carrasco-Salas et al., 2019; Ginno et al., 2012). We verified that this sequence, when transcribed in *S. cerevisiae*, generates high levels of RNase H-sensitive R-loops, as determined by DRIP followed by quantitative PCR (Figure S3B). Prominent H-CRAC signals were detected at the ectopically expressed mAIRN *locus* (Figure S3C), validating the notion that H-CRAC robustly identifies well-established regions of R-loop formation.

H-CRAC signals overlapped transcription at the genome-wide scale, as expected considering the co-transcriptional nature of R-loop formation (Figures 4A, S3D and S3E), but the pattern of the RNAPII and RNase H signals was generally different (Figures 4A and S3E), consistent with the notion that not all transcribed regions generate R-loops to the same extent.

H-CRAC signals with both RNase H1 and RNase H2 were markedly strand specific (Figure S3E), with distinct and well-defined peaks (Figure 4A). A comparison with the only directional R-loop map generated by S9.6-DRIP-seq in yeast (San Martin-Alonso et al., 2021) revealed significant overlaps when taking into account the overall signal along coding genes (Figures 4A, 4B). However, the resolution of the signal was clearly higher and its distribution within genes often different, with prominent H-CRAC signals observed at new locations (Figure 4A and data not shown). The detailed genome-wide analysis of R-loops distribution determined by H-CRAC is beyond the scope of this report and will be provided in a separate manuscript (Aiello et al., in preparation).

Sensitivity of signals to overexpression of human RNase H is a landmark to gauge the reliability of R-loop detection (Chédin et al., 2021). Therefore, we overexpressed human RNH1 in cells that also expressed a version of Rnh201 suitably tagged for H-CRAC, and analysed the data after normalization to an *S. pombe* spike in. Rnh201 H-CRAC signals were considerably reduced in many locations (Figure 4C), without a significant, general effect on transcription (Figure 4C and S3F), which was monitored in parallel to ascertain that reduced H-CRAC signals were not due to altered gene expression. This indicates that although ectopically expressed hsRNH1 cannot fully outcompete tagged endogenous Rnh201, it can significantly reduce R-loop levels. Consistently, we could also generate similar H-CRAC signal distributions using tagged hsRNH1, confirming that the ectopically expressed, heterologous enzyme recognizes very similar targets as the yeast proteins (Figures S3G and S3H). Together, these data demonstrate that H-CRAC is a sensitive and resolutive method for detecting *in vivo* at least a significant fraction of cellular DNA:RNA hybrids, overlapping and complementing *in vitro* DRIP-based methods. Similar to DRIP, H-CRAC cannot distinguish between DNA:RNA hybrids and R-loops, the latter being defined by the presence of a single stranded DNA portion. However, known non-R-loop DNA:RNA hybrids would either be too short (e.g., Okazaki fragments) or protected from RNase H crosslinking (e.g., the hybrid contained in the RNAPII inner channel) to be detected by H-CRAC. With this caveat in mind, we will therefore use preferentially the terms R-loops or RNase H targets for the structures generating H-CRAC signals.

To assess whether R-loops form at the rRFB during replication, we generated H-CRAC maps in *sen1-3* and wild-type cells. Interestingly, we detected prominent Rnh1 H-CRAC signals peaking roughly 50 nt upstream of the fork-trailing RNAPII peak (Figures 3B and 4D), consistent with the notion that they form immediately upstream of the stalled polymerase. Rnh201 signals were less prominent in this position and were instead preferentially observed roughly 500 nt upstream (Figure S4A), suggesting that Rnh1 might preferentially recognize R-loops at this site. Ectopically-expressed human RNase H (hsRNH1) for H-CRAC generated

clear signals that nicely overlapped yeast Rnh1 targets in this region (Figure 4D), further supporting the notion that they represent *bona fide* R-loops.

H-CRAC signals proximal to the rRFB increased significantly in *sen1-3* cells in their absolute levels but not when evaluated relative to the levels of paused RNAPII upstream of the replication fork (Figure 4D). This finding indicates that higher R-loop levels at this site parallel the increased stalling of RNAPII engaged in conflicts with replication forks when Sen1 cannot interact with the replisome.

From these experiments we conclude that R-loops form upstream of RNAPII conflicting with the stalled replication forks. The interaction of Sen1 with the replisome is required for releasing fork-trailing RNA polymerases, but not for limiting the levels of R-loops that form *per* transcriptional event.

RNases H promote RNAPII release at the rRFB

The strong growth defect of *sen1-3* cells in the absence of RNase H activity, suggests that Sen1 at the replisome and RNases H have a common or complementary function either in limiting R-loops accumulation, in RNAPII release, or both. Because we did not observe increased R-loops *per* transcription event at the rRFB in *sen1-3* cells, we considered a possible implication of RNases H/R-loops in RNAPII release. We first monitored RNAPII occupancy by CRAC in a *rnh1Δ rnh201Δ* mutant, which lacks RNase H activity. In this genetic context, transcription was not generally altered, as shown by the profiles of median RNAPII CRAC signals on genes (Figure S4B). Interestingly, however, we observed a clear increase in the levels of RNAPII at the rRFB compared to a wild-type strain, which accumulated in the same position and to similar levels as in *sen1-3* cells (Figure 4D). Thus, RNases H are required to limit accumulation of RNAPII at the TRC in the rRFB.

RNases H might contribute to the release of the polymerase independently of R-loop degradation, or by degrading the RNA moiety of these structures. In this latter perspective the expected increase in R-loops in *rnh1Δ rnh201Δ*, might prevent efficient RNAPII release. One important prediction of this hypothesis is that degrading the R-loops formed upstream of the

stalled polymerase in *sen1-3* cells should favour its release and suppress the RNAPII accumulation phenotype.

Overexpression of human hsRNH1 significantly reduced R-loop levels genome-wide, without generally altering transcription, as determined by RNAPII CRAC analyses (Figures 4C and S3F), yet this was accompanied by a significant reduction in RNAPII accumulation at the rRFB in *sen1-3*, to similar levels as in wild-type cells (Figure 4D). Together, these results suggest that modulating the levels of R-loops at the rRFB, either by increasing them in *rnh1Δ rnh201Δ* cells or by decreasing them upon overexpression of hsRNH1, affects RNAPII release in an anti-correlative manner.

In the light of these results, we considered that RNases H might also contribute to the release of RNAPII that stall at TRCs in the 5'-end of genes undergoing replication. In this perspective it is expected that the absence of RNase H activity should generate a 5'-skewed RNAPII pattern similar to the one observed in *sen1-3* cells in asynchronous cells (Figure 1A). However, the double *rnh1Δ rnh201Δ* mutant did not phenocopy the *sen1-3* RNAPII 5'-end accumulation suggesting that RNases H do not partake in releasing RNAPII in these locations (Figure S4B). This conclusion is also supported by the findings that overexpression of hsRNH1 in *sen1*-3 cells did not suppress 5'-end RNAPII accumulation (Figure S4C, compare with Figure 2A) and that R-loops are not significantly detected in the very 5'-end of genes where the increased RNAPII accumulation is observed (Figures S4D and S4E). One likely explanation for these results is that in the proximity of the TSS the short length of the available nascent RNA is not compatible with the formation of DNA:RNA hybrids (see Discussion).

From these experiments we conclude that RNAPII is released at the rRFB by the combined action of Sen1 at the replisome and RNase H, which presumably exerts its function by degrading the R-loops formed in the wake of the stalled transcription elongation complex.

Sen1 releases RNAPII at sites of conflicts with RNAPIII

Aside from RNAPII and replisome components, Sen1 also interacts with RNAPIII, an interaction that is also lost *in sen1-3* cells (Xie et al., 2021). We have shown that this latter

interaction is not mediated by the replisome but reflects the existence of an alternative complex since quantitative MS analysis of the RNAPIII interactome detected Sen1 but not replisome components.

We have previously shown that non-coding and non-annotated RNAPII transcription events might collide with RNAPIII transcription units, where they are restrained by roadblocks located in the 5'- and 3'-end of tRNA genes (Candelli et al., 2018). The mechanism by which RNAPII is released at these sites was not addressed, although we showed that RNAPII ubiquitylation and degradation occurred at other sites of roadblock. We hypothesized that, analogous to the role at the replisome, interaction with RNAPIII might recruit Sen1 for removing conflicting RNAPIIs. Therefore, we monitored RNAPII transcription around tRNA genes in wild-type and *sen1-3* cells. RNAPII was found to accumulate upstream of tRNA genes in *sen1-3* cells, consistent with the notion that conflicting RNAPII is not released efficiently when Sen1 cannot interact with RNAPIII. This can be clearly appreciated at individual cases (Figure 5A), and more generally by the heatmap analysis of the whole tRNA genes population (Figure 5B). We considered the possibility that this accumulation was related to ongoing replication, but the same analysis performed in G_1-arrested cells revealed very similar effects (Figures 5A and 5B). Increased accumulation of RNAPII was also observed when monitoring antisense transcription relative to tRNA genes (Figures 5A and 5B, right panels), although to lower levels, possibly because additional mechanisms are in place for limiting head-on conflicts. Nevertheless, we clearly observed that, when uncontrolled by Sen1, RNAPII entered tRNA transcription units in the antisense direction (for technical reasons we cannot monitor sense RNAPII transcription in the body of tRNA genes). Similar to what observed at mRNA coding genes, no significant effect was observed after deletion of the two RNase H genes (Figure S5A), strongly suggesting that degradation of R-loops is not playing an important role at these sites.

From these observations, we conclude that Sen1 plays important roles in limiting conflicts between RNAPII and RNAPIII at tRNA genes, a role that is dependent on its interaction with RNAPIII but is independent from ongoing replication.

Figure 5

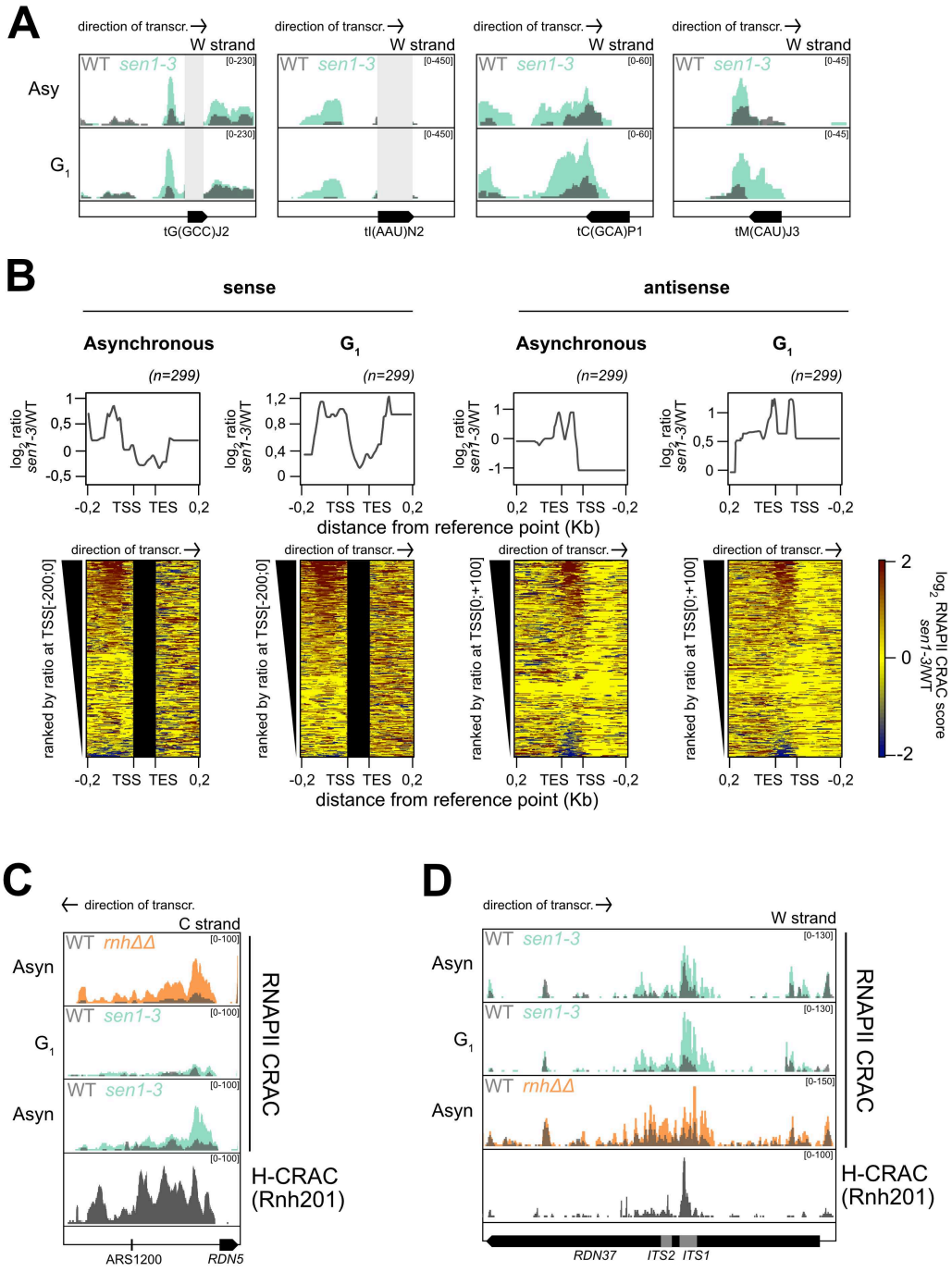


Figure 5: Sen1 releases RNAPII at sites of conflicts with RNAPIII and limits RNAPII transcription in the ribosomal DNA

A) RNAPII accumulation in the sense (left panels) or antisense (right panels) orientation at individual, representative tRNA genes in *sen1-3* cells. WT and *sen1-3* signals have been overlapped for easy comparison as in Figure 2. The strand of the tRNA gene and the direction of RNAPII transcription are indicated. Sense RNAPII signals within the tRNA body are masked because they cannot be reliably distinguished from contaminating mature tRNAs. Asyn: asynchronously growing cells; G_1: cells arrested in G_1 with α-factor.

B) Heatmap analyses representing the $\log_2$ ratio of the RNAPII signal in *sen1-3* vs WT cells at tRNA genes aligned on their TSS and TES respectively for the sense (left panels) and the antisense transcription (right panel). For RNAPII transcription that is sense relative to the tRNAs, genes were ranked by the $\log_2$ ratio in the 100 nt window preceding the TSS, which is where the RNAPII peak is generally observed. As in A, signals within the tRNA body are masked. For antisense transcription, it is possible to monitor antisense RNAPII signals within the body of the tRNA gene, which is where the ratio *sen1-3*/WT increases the most. In these heatmaps, ranking was done based on this region. The summary plot on the top was calculated using the median values for each position.

C) RNAPII occupancy antisense and upstream of the *RDN5* gene, in *sen1-3* and RNases H deleted cells (***rnhΔΔ***). H-CRAC signal for Rnh201 is also shown for the same region**.** The C strand is monitored (antisense of *RDN5* transcription) and the direction of RNAPII transcription is indicated. **D)** CRAC RNAPII occupancy antisense to the *RDN37* gene in WT, *sen1-3* and RNases H deleted cells (***rnhΔΔ***). R-loops detection by Rnh201 H-CRAC is also shown.

Roles of Sen1 and RNases H in limiting RNAPII transcription in the ribosomal DNA

Although the ribosomal *loci* are mainly devoted to the production of rRNA by RNAPI and RNAPIII, transcription by RNAPII is also known to occur, mainly antisense to the transcription unit producing the 37S rRNA precursor. Because the 5S rRNA is produced by RNAPIII, we first investigated whether the *sen1-3* mutation would affect RNAPII occupancy around the *RDN5* gene. Akin to other RNAPIII transcription units, we clearly observed replication-independent (i.e., observed also in G_1-arrested cells) RNAPII accumulation antisense of *RDN5* in *sen1-3* cells (Figure S5B). RNAPII accumulation was also observed in ***rnh1Δ rnh201Δ*** cells suggesting that RNase H is required at this *locus* for efficient RNAPII release. Consistently, R-loops were detected, both by H-CRAC and DRIP (San Martin-Alonso et al., 2021; Wahba et al., 2016) immediately upstream of the paused polymerase (Figure S5B).

Interestingly, we also noticed the occurrence of transcription upstream and antisense of *RDN5* directed towards the close rARS (*ARS1200*) replication origin (Figure 5C), which was

only observed in replicating cells and was also RNases H dependent. R-loops were detected by H-CRAC overlapping the whole region of transcription separating the replication origin from the *RDN5* gene, most likely landmarking sites of head-on transcription-replication conflicts. Thus, around *RDN5* Sen1 has replication-dependent and -independent roles in limiting conflicts involving RNAPII transcription.

These findings prompted a closer examination of RNAPII transcription in the rDNA repeats in *sen1-3* and *rnh1Δ rnh201Δ* mutants. Interestingly, we also found in both mutants a region of increased RNAPII occupancy antisense of *RDN37* transcription, roughly corresponding to the internal transcribed spacer 1 (*ITS1*, Figure 5D). This increase was not dependent on replication as it was observed in G_1-arrested cells and was associated to the formation of R-loops detected with H-CRAC for both Rnh1 and Rnh2 (see Discussion).

We conclude from these data that Sen1 and RNases H play important roles in resolving conflicts involving RNAPII transcription in the ribosomal DNA region.

RNase H activity and the dual roles of Sen1 transcription termination and in conflict-solving are required for maintaining genome stability

We set up to explore the consequences of the *sen1-3* mutation on genomic stability. We first assessed the sensitivity of *sen1-3* cells to replication stress induced by hydroxyurea (HU) and methyl methanesulfonate (MMS). We found a moderate and strong sensitivity to HU and MMS respectively (Figure 6A), indicating that association of Sen1 with the replisome is required for optimal replicative stress response. However, and as we had previously described (Appanah et al., 2020) we did not observe the same genomic instability phenotypes reported for the loss-of-function *sen1-1* mutant, both at the level of Transcription Associated Recombination (Mischo et al., 2011) (Figure S6A), and Rad52 *foci* formation, a hallmark of DSB accumulation and repair (see below, Figure 6D). Also, we did not observe an increase in R-loops by H-CRAC (Figures S4E, S4F and S4G), coherent with previous immunofluorescence analysis of chromosome spreads (Appanah et al., 2020). These differences could be explained if the alterations in the transcription landscape observed in

Sen1 loss-of-function but not in *sen1-3* cells (Figure 1) contribute significantly to the genomic instability phenotypes. Also, the functional cooperation with RNases H, underlying the strong genetic interaction with *sen1-3,* might mask phenotypes linked to the loss of interaction with the replisome and RNAPIII. Therefore, we devised a genetic system for analysing the damage induced by the combination of the *sen1-3* and *rnh1Δ rnh201Δ* mutations. We constructed an inducible triple mutant whereby a chromosomal wild-type, AID-tagged Sen1 complements lethality of a *rnh1Δ rnh201Δ* strain containing a plasmid-borne, untagged *sen1-3* allele. Upon addition of auxin (IAA) the wild-type copy is rapidly degraded leaving only the Sen1-3 protein (Mendoza-Ochoa et al., 2019). Induced triple mutant cells only partially recapitulated the lethality of the *bona fide* triple mutant, possibly because wild-type Sen1 was not fully depleted and/or slight overexpression of *sen1-3* partially suppressed its phenotype (Figure S6B). However, combination of the IAA treatment and incubation at the suboptimal temperature of 37°C led to a similar level of growth impairment as the one observed for Sen1 depletion. We thus carried our DNA damage analysis also at this temperature. We first monitored the extent of DNA damage by measuring the frequency of Rad52-YFP *foci* upon induction of the triple mutant phenotype (Figures 6B, S6C and S6D). In the non-induced triple mutant, we observed, as expected, the same Rad52 *foci* frequency observed in a double *rnh1Δ rnh201Δ* deletion. Importantly, partial induction of the triple mutant phenotype led to a significant increase in the number of Rad52-YFP *foci* compared to the double RNases H deletion, implying that in this context association of Sen1 with the replisome is required for limiting DNA damage.

Another hallmark of DNA damage, H2A histone phosphorylation at position S129, was monitored by western blot detection. Consistent with the increased Rad52-YFP *foci*, phosphorylation of H2A was found to be significantly increased when the triple mutant, partial phenotype was transiently induced by the addition of IAA (Figures 6C and S6E).

To assess if the transcription termination defects in Sen1 loss-of-function mutants significantly aggravate the phenotypes linked to the absence of Sen1 at the replisome, we induced a non-coding transcription termination defect in *sen1*-3 independently of Sen1 by additionally mutating Nrd1, another component of the NNS complex. We reasoned that if the

Sen1 loss-of-function phenotypes were due to a combined termination and conflict-solving defect, a double *nrd1-102 sen1-3* mutant should recapitulate the phenotypes of *sen1-1* cells. This expectation was fully met when analysing the cellular frequency of Rad52 *foci* in the single and double mutants. *Foci* were poorly detected in single mutants, consistently with previous reports (Appanah et al., 2020; Costantino and Koshland, 2018; Mischo et al., 2011), but were found to similar levels in *sen1-1* and *nrd1-102 sen1-3* cells (Figure 6D).

From these results we conclude that by interacting with the replisome and RNAPIII, Sen1 plays important roles at sites of transcription-transcription and transcription-replication conflicts. In the rDNA this role is redundantly exerted by RNases H. In the absence of both enzymes or in the presence of transcription termination defects, extensive DNA damage occurs, which is likely underlying lethality.

Figure 6

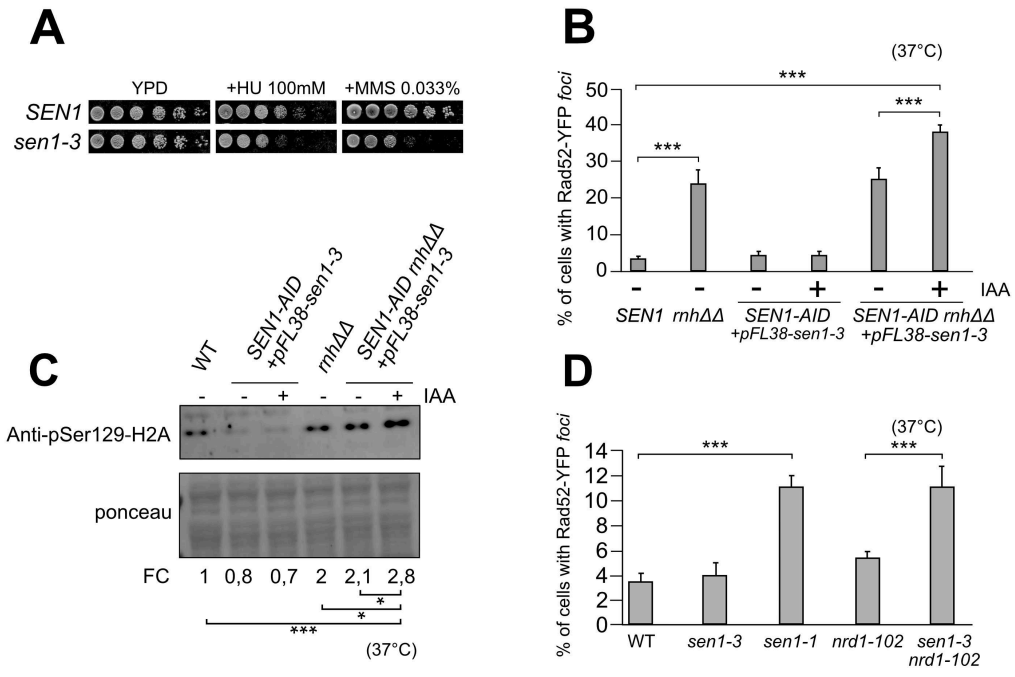


Figure 6: Sen1 cooperates with RNases H to maintain genome stability

A) Growth assay of *sen1-3* and WT cells as a control in the presence or absence of hydroxyurea (HU) and methyl methanesulfonate (MMS) as indicated. Plates were incubated for 3 days at 30°C. Growth assays were performed

on the same plates for each series (i.e., HU or MMS). **B)** Frequency of cells containing Rad52-YFP *foci* in asynchronously growing cultures incubated at 37°C for 1 h for the indicated strains, in presence or absence of auxin (IAA) to deplete WT Sen1-AID. *** p<0.001. Standard deviation (n=3) is indicated. **C)** Western blot detection of H2A Ser129 phosphorylation in asynchronously growing cultures incubated at 37°C for 1 h for the indicated strains, in presence or absence of auxin (IAA) to deplete WT Sen1-AID. The average level quantified from three independent replicate (Figure S6C) is shown at the bottom. * p<0.5; *** p<0.001. **D)** Rad52-YFP *foci* are monitored in the indicated strains as in Figure 6B. Coupling of the *sen1-3* allele to Sen1-independent transcription termination defects generated by the *nrd1-102* leads to levels of DNA damage comparable to the ones observed in *sen1-1* cells. *** p<0.001. Standard deviation (n=3) is indicated.

DISCUSSION

Synchronous occupancy of the compact yeast genome by several cellular machineries that exert essential functions in the expression, maintenance and transmission of the genetic information requires the existence of robust mechanisms that ensure the coordination of the different processes, prevent conflicts, or solve them when they occur. Dealing with the extensive occupancy of the genome by transcription events that largely overcome the limits of functional gene annotations is a major challenge in this context. In this study we employed high resolution and innovative genomic tools to elucidate the functions of the Sen1 helicase and RNases H in “genomic distancing”. Importantly, and as opposed to previous reports (Alzu et al., 2012; Costantino and Koshland, 2018; Mischo et al., 2011; Zardoni et al., 2021), this study was performed without altering the physiological transcriptional landscape and in the absence of global replication stress. We propose a model that implicates these enzymes in many sites to control transcription-replication and transcription-transcription conflicts. Importantly, we revisit the role of Sen1 in genomic stability and transcriptional homeostasis by disentangling these functions, whose synthetic association likely impinges on the many phenotypes previously described.

Loss of Sen1 function generates termination and conflict-solving defects that conjunctly lead to genomic instability

Many earlier studies, including from our laboratory, have clearly established a role for Sen1 in transcription termination of several thousands of non-coding RNA genes (for a review see Porrua et al., 2016). The biochemical mechanism of Sen1-dependent termination has also been extensively studied *in vitro* (Porrua and Libri, 2013; Han et al., 2017; Leonaitė et al., 2017; Wang et al., 2019). Genome-wide analyses of the effects of Sen1 depletion (this study; see also Schaughency et al., 2014) demonstrate the inductions of major alterations in the transcriptional landscape. These alterations in the transcription scenery have a large potential for increased interference with concurrent processes, together with global alterations in gene expression homeostasis and cellular physiology. Indeed, we also describe changes in the

gene expression program, which, at least to some extent, parallel the ribosome assembly stress described recently by the Shore and Churchman laboratories (Tye et al., 2019; Zencir et al., 2020). These phenotypes are possibly triggered by defective production of snoRNAs, which are major NNS targets and contribute to rRNA maturation, or by defects in transcription termination of rRNA genes, a process in which Sen1 has been implicated (Kawauchi et al., 2008). Together, these results demonstrate that the transcription and gene expression program of Sen1 loss-of-function mutants is largely distinct from the one of wild-type cells, and raise questions about the assessment of genomic instability phenotypes in such non-physiological conditions.

It has been shown that alterations in transcription termination generated with other NNS mutants (i.e., *nrd1* and *nab3*) do not induce instability *per se* (Alzu et al., 2012; Costantino and Koshland, 2018; Mischo et al., 2011). These earlier findings indicate that transcription termination defects are not sufficient, alone, to generate genomic instability, most likely because control or prevention of conflicts rely on robust and redundant mechanisms. Still, direct or indirect effects of impaired termination can generate instability when coupled to other defects due to altered Sen1 functions. Fully consistent with this notion is our finding that impairing transcription termination independently of Sen1 (i.e., by mutation of Nrd1) in *sen1-3* cells, fully recapitulates the DNA damage phenotype of *sen1-1* cells, which are defective for both termination and conflict-solving functions of Sen1 (Figure 6). The simplest interpretation of these results is that when termination alone is impaired, the increased transcriptional challenge at sites of conflicts can still be resolved by Sen1 and RNases H, while in the additional absence of this conflict-resolving function increased DNA damage occurs or is not repaired efficiently.

On the physiological relevance of Sen1 binding to the replisome

Focal to our study is the *sen1-3* mutant, which we have previously shown to fully lose interaction with replisome components (Appanah et al., 2020) and with RNAPIII (Xie et al., 2021). Our working hypothesis underlying this study has been that the interaction of Sen1 with

the replisome allows either its recruitment or its function at sites where transcription collides with replication to dismantle the elongation complex for giving way to replication. We have shown that *sen1-3* cells have no transcription termination defects or altered gene expression, which allowed studying the role of Sen1 at sites of conflicts under overall physiological conditions in terms of transcription. Because *sen1-3* cells have no major growth or hyper-recombination phenotypes (Figures 6 and S6), it would be legitimate to question the physiological relevance of the functions impaired in *sen1-3* cells. However, several observations support the notion that the *sen1-3* mutation affects important facets of Sen1 function: i) viability is not supported in the absence of RNase H activity (Figure S6B); ii) growth is affected in the presence of genotoxic agents (Figure 6A); iii) the speed of the replication forks is decreased (Figure S2H); iv) DNA damage levels increase in a *rnh1Δ rnh201Δ* background or when transcription termination is additionally affected (Figures 6B and 6D); v) altered RNAPII distribution is detected in several locations where TRCs are expected to occur (Figure 2); vi) defective RNAPII release occurs at sites of conflicts with RNAPIII (Figure 5). These findings underscore the notion that redundant mechanisms are in place to back up Sen1 action and/or that major effects can only be observed under challenging conditions.

In this perspective, while this work was in progress it was reported that DNA damage and fork stalling occur under Sen1 depletion at sites of TRCs specifically under conditions of HU-induced replication stress (Zardoni et al., 2021). In this study, transcription of genes that host TRCs was also found to be altered, a phenotype that we did not observe even upon Sen1 depletion (data not shown). Our interpretation for these differences is that failure to complete transcription at some genes hosting TRC sites results from a synergistic effect of decreased replication fork progression due to the HU treatment and the failure to efficiently remove RNAPII by Sen1 at TRCs, which together delays the resolution of conflicts and affect the resumption of novel transcription cycles.

We provide evidence that Sen1, by binding to the replisome, releases RNAPII engaged in conflicts with replication in the very 5'-end of genes, which is the region with the highest average RNAPII persistence in yeast. We also found increased *sen1-3*-dependent RNAPII

accumulation in other regions of pausing (Figure 2B) suggesting that it is RNAPII pausing and not the 5'-end of genes *per se,* that determines the preferential sites of conflicts. A similar, genic RNAPII accumulation was described in *dicer* mutants in the 3'-end of *S. pombe* genes, which was also proposed to be due to the defective resolution of TRCs (Castel et al., 2014). The reasons for these differences in the position where RNAPII accumulates are not clear, but they might be due to a different mechanism of action of Sen1 and Dicer or to the different distribution of RNAPII pausing at *S. pombe* genes.

While this work was in preparation, a similar increase in RNAPII occupancy was reported to occur in cells expressing loss-of-function mutants of the human Sen1 homologue *SETX* or in Δ*SETX* cells (Kanagaraj et al., 2022). Although it was not shown whether these phenotypes are due to TRCs, they were found more prominently in very long (>100kb) genes, which generally host common fragile sites (CFS). Genes hosting CFS were also shown in this study to be frequently subject to genomic rearrangements in the absence of SETX. Thus, the role of Sen1 in solving TRCs within genes might be conserved for human SETX, and could also be independent from its function in terminating non-coding transcription, which is not clearly established in human cells.

On the relationship between Sen1 and R-loops

A salient question, central to this study, concerns the functional relationships between Sen1 and RNases H. Although both classes of enzymes have been involved in the resolution of R-loops and have been proposed to work together at these structures (Costantino and Koshland, 2018), we could not globally detect increased R-loop levels in *sen1-3* cells by H-CRAC (Figures S4E-G), chromosome spreads (Appanah et al., 2020), or by DRIP qPCR at a few R-loop-prone sites (data not shown). This indicates that the interaction of Sen1 with the replisome is not required for suppressing R-loops and might suggest that the R-loop increase detected in *sen1* lack-of-function backgrounds is also linked to the transcriptional termination defect of this mutant background. Because R-loops are known to be sites prone to DNA damage, this notion is consistent with the finding reported here that increased sites of DNA

damage are observed when associating a termination defect generated by the *nrd1-102* allele to the *sen1-3* mutation. In this perspective, it is possible that Sen1, rather than unwinding R-loops that are constitutively formed, prevents their formation by a dual action in restricting non-coding transcription and solving conflicts with replication. In agreement with this hypothesis, we have previously reported biochemical and single-molecule evidence that purified Sen1 is poorly processive in unwinding DNA:RNA duplexes (Porrua and Libri, 2013; Wang et al., 2019).

Sen1 and RNases H cooperate to release RNAPII at the ribosomal DNA

At the rRFB and upstream of *RDN5,* both the *sen1-3* and the *rnh1Δ rnh201Δ* mutations lead to replication-dependent, increased RNAPII occupancy, suggesting that both factors play complementary or redundant functions in the efficient removal of transcription complexes. Transcription upstream of the rRFB was previously shown to be required for inducing rDNA copy number amplification. It was shown that bi-directional transcription generated from the E-pro region using the strong *GAL10* promoter releases cohesin complexes and induces rDNA amplification in a strain containing only two rDNA repeats (Kobayashi and Ganley, 2005). Because of its stability and possibly its size, it is unlikely that the transcript we describe here corresponds to the E-prom transcripts reported by Kobayashi et al. (Kobayashi and Ganley, 2005). However, it is possible that Sen1 and RNases H provide an additional layer of control preventing RNAPII transcription from invading the rDNA, which could cause genomic instability in the region. It is also possible that efficiently removing RNAPII from upstream of the stalled fork is important to limit replication stress and favor the progression of replication in this region where two replication forks converge.

A direct role of RNases H in termination could complement the function of Sen1 when recruited by the replisome, and possibly underlie the synthetic lethality observed. In a mechanistic perspective it can be envisioned that the removal of R-loops by RNases H weakens the stability of the elongation complex, facilitating its dismantling at sites of conflicts. For instance, this could occur if, upon pausing, RNAPII would backtrack on a substrate on

which RNase H has previously removed the DNA-associated nascent transcript engaged in an R-loop. Alternatively, it is possible that RNases H might favour RNAPII release at the rRFB (and possibly other sites) by cleaving the R-loop-engaged RNA thus providing an entry site for the Rat1 exonuclease, which would degrade the 3' portion of the nascent RNA and terminate transcription by the "torpedo" mechanism (Porrua et al., 2016). If Rat1 is a downstream effector of RNases H in at least some genomic locations, it could be expected that its mutation strongly affects growth of *sen1-3* cells, phenocopying the double *rnh1Δ rnh201Δ* deletion. Indeed, associating the thermosensitive *rat1-1* allele to the *sen1-3* mutation induced a strong synthetic growth defect at the semi-permissive temperature for *rat1-1* (Figure S4H, compare growth of *rat1-1* to *rat1-1 sen1-3* cells). This phenotype is unlikely due to defective termination of mRNA-coding genes in *rat1-1* cells because the even stronger defect induced by mutation of Rna15, an essential termination factor, did not generate a similar synthetic phenotype when associated to *sen1-3* (Figure S4H, compare growth of *rna15-2* to *rna15-2 sen1-3* cells). Interestingly, a role in transcription termination for RNase H is in agreement with a recent report showing that cleavage of the nascent RNA by oligonucleotide-directed digestion (possibly mimicking R-loop digestion) could induce torpedo-generated transcription termination *in vivo* (Lai et al., 2020). Finally, it is also possible that digestion of the DNA:RNA heteroduplex by RNase H provides a better opportunity for Sen1 to efficiently access the nascent RNA close to the stalled elongation complex and induce termination. Indeed, we have previously shown that *in vitro* Sen1 cannot access DNA:RNA or RNA:RNA double stranded regions of the nascent transcript, the presence of which actually hampers Sen1-dependent termination when close to the RNAPII (Porrua and Libri, 2013; Xie et al., 2021). Besides reducing its binding opportunities, R-loops might also hinder the translocation on the nascent transcript of Sen1, which we have shown to have *in vitro* poor processivity when unwinding heteroduplexes (Porrua and Libri, 2013; Wang et al., 2019).

Sen1 resolves transcription-transcription conflicts

We demonstrate that mutation of Sen1 has NNS termination-independent effects in many genomic regions (Figure 7), but with different modalities. In *sen1-3* cells, accumulation of RNAPII at the rRFB and upstream of *RDN5* is dependent on ongoing replication, is associated to the formation of R-loops and is also dependent on RNases H.

Figure 7

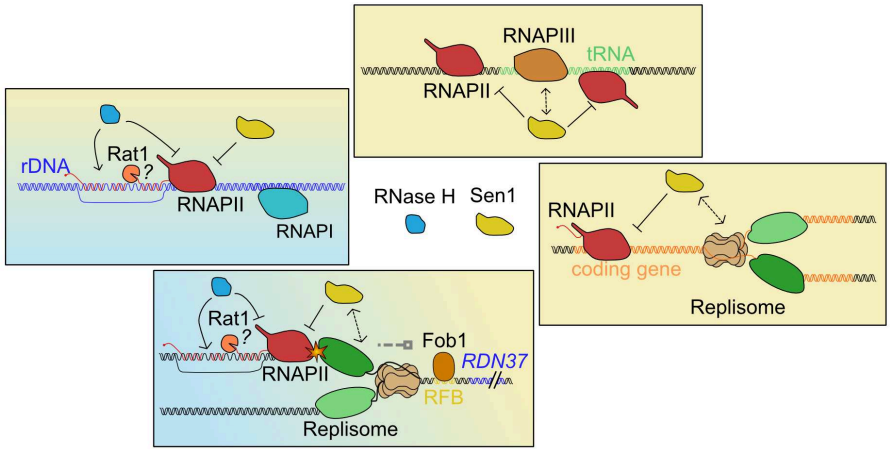


Figure 7: Model of the function of Sen1 and RNases H in controlling transcription replication and transcription-transcription conflicts.

Sen1 resolves transcription-replication and transcription-transcription conflicts by dislodging RNAPII from several locations including the 5'-end of coding genes, the rRFB, the rDNA and tRNA genes. During TRCs Sen1 is recruited via interaction with the replisome, while at tRNA genes the recruitment occurs *via* the interaction with RNAPIII. RNases H assist Sen1 in limiting RNAPII at the rRFB and in the rDNA, via a mechanism that could involve Rat1.

Conversely, we did not observe a role for RNases H, nor an effect of hsRNH1 overexpression in limiting replication-dependent RNAPII accumulation at genes, suggesting that R-loops are not formed in these TSS-proximal locations. The peak of differential RNAPII accumulation in *sen1-3* cells is located roughly 100 nt after the TSS (Figure 2A). Considering the physical occupancy of the polymerase, the binding of the capping enzymes and the

capping complex, it is possible that the limited window of free nascent RNA does not allow efficient formation of R-loops.

At tRNA genes, although R-loops were detected at some genes upstream of paused polymerases (data not shown) RNases H do not appear to play a significant role, maybe because alternative mechanisms are in place cooperating with Sen1 to release the polymerase as previously shown for other sites of roadblock (Candelli et al., 2018; Colin et al., 2014). In these cases, Sen1 is solving conflicts between RNAPII and RNAPIII, by virtue of its interactions with the latter that is also lost in *sen1-3*, and, consistently, RNAPII accumulation is observed also in the absence of replication. The fact that the same region of Sen1 mediates alternative interactions with RNAPIII and the replisome might have important functional implications, to ensure that the two functions of the helicase, although mechanistically similar, remain distinct, and that replisome components and RNAPIII are never found, inappropriately, in the same complex, connected by Sen1. Perhaps this region of Sen1 mediates contacts with other molecular machineries to exert similar functions. In this regard, we observed marked RNAPII persistence in *sen1-3* cells antisense of the *RDN37* transcription unit, in correspondence of the *ITS1*. Limiting RNAPII in this region is also dependent on RNases H and prominent levels of R-loops are observed by H-CRAC. However, this accumulation was also observed in the absence of replication, and why it occurs in the *sen1-3* mutant is presently unclear. One possibility is that the interaction with another factor, responsible for recruiting Sen1 at this site, is also lost in the *sen1-3* mutant and that RNAPII persists at sites of head-on conflicts with RNAPI. It is enticing to speculate that Sen1 is recruited to the nucleolar site of rRNA transcription by RNAPIII, while transcribing the 5S rRNA or tRNAs that are also transcribed in clustered nucleolar regions (Thompson et al., 2003; Haeusler and Engelke, 2006). Loss of Sen1 from RNAPIII complex in *sen1-3* cells would also bring about defective management of RNAPI-RNAPII conflicts.

In the light of the impact of the *sen1-3* mutation on RNAPIII termination, we considered the possibility that the accumulation of RNAPII at the rRFB might be linked to its role at the upstream *RDN5* gene. However, RNAPII accumulation occurs in a position that is clearly

downstream of the region of RNAPIII readthrough, as shown by RNAPIII CRAC analyses (Figure S4A).

Together, these results, obtained in the absence of possibly interfering transcriptional defects, allow attributing to Sen1 a role of “master conflicts regulator” that is similar in many aspects to the one described for Dicer in *S. pombe* (Castel et al., 2014). Dicer was indeed implicated in releasing RNAPII at genes (in this case in their 3’-end), antisense of tRNA genes and rRNA transcription and at sites of replication stress. Although the mechanism of action of the two factors is unlikely to be similar, converging evolution might have hijacked existing cellular mechanism to fulfill the important role of coordinating essential cellular processes.

H-CRAC is a suitable method for R-loops detection genome-wide

We describe here a novel method to detect R-loops with unprecedented sensitivity and resolution. H-CRAC meets all the essential landmark requirements we assessed for bona fide R-loop detection (Chédin et al., 2021) strongly supporting the notion that RNase H targets detected by this method represent *bona fide* R-loops. It is important to stress that H-CRAC is fundamentally different from ChIP as it detects the interaction of RNases H with the RNA and is not expected to sense recruitment of the enzyme to the DNA in the absence of a specific contact with its targets.

Our maps are similar to published DRIP-seq maps, as witnessed by the statistically significant overlap for the R-loop-forming genes detected by H-CRAC and the only directional map available (San Martin-Alonso et al., 2021) (Figure 4B). Comparisons with other published data (Achar et al., 2020; Wahba et al., 2016), are also statistically significant (data not shown) but are less reliable, considering the non-directional nature of these studies. Nevertheless, in many cases differences are also observed between H-CRAC and DRIP-seq (Figure 4A). Some of these differences can be accounted for by the better resolution and most likely higher sensitivity (i.e. the signal to background ratio) of H-CRAC relative to DRIP-seq (Figure 4A). However, it is also possible that the targets detected by the two methods are, to some extent, inherently different. It was suggested that short R-loops might not survive the extraction and

immunoprecipitation steps of DRIP (Chédin et al., 2021), still they could be detected by *in vivo* crosslinking to RNases H. Conversely, some stable R-loops detected by DRIP might not be efficiently recognized by RNases H. Thus, the two methods might provide overlapping and complementary outputs for a better understanding of the distribution, metabolism and functional implications of DNA:RNA hybrids. The thorough analyses of R-loop distribution in yeast, the differences between Rnh1 and Rnh201 and the relationships with other detection methods are beyond the scope of this report and will be detailed elsewhere. However, to the light of the results and controls presented here we trust that H-CRAC will provide invaluable information to study R-loop biology and the relationships with transcription and genome maintenance.

MATERIALS AND METHODS

Yeast strains, plasmids and oligonucleotides

The strains, plasmids and oligonucleotides used in this study are listed in Key Resource Table. Yeast strains used in this study derive from W303 or BMA64, which is a *trp1*Δ derivative of W303. Strains newly modified were constructed with standard procedures (Longtine et al., 1998).

Transcription-Associated Recombination (TAR) assay

To assess the frequency of recombination, strains of interest were transformed with the pRS314-L recombination reporter at 30°C. Recombination events were scored by assessing the number of cells containing a functional, recombined *LEU2* gene relative to the total number of cells plated. Six colonies of at least three independent transformants were analysed.

Cell growth for CRAC and Copy Number experiments

For each condition, 2 L of cells expressing an HTP-tagged version of the protein of interest expressed either from the endogenous locus (i.e., Rpb1, Rnh1, Rnh201) or from a plasmid (hsRNH1) were grown in logarithmic phase to OD_{600}=0.6 at 30°C in a CSM-TRP medium. Cells ectopically expressing the mAIRN construct were grown in CSM-Trp-Ura. Cells over-expressing hsRNH1 were grown in CSM-Trp-His.

G_1 cell cycle arrest was triggered at OD_{600}=0.3 by 3 consecutive additions of 4, 8 and 4 mg of α-factor spaced by 40 min. 40 min after the last addition of α-factor, and before UV-crosslinking, G_1 arrest was verified both by microscopic visualisation of cell morphology and by flow cytometry (Figure S2J).

For analyses in S-phase, cells were arrested in G_1 by α-factor as described above and released into S-phase by removing α factor by filtration on a glass microfiber filter (pore ⌀=1.6 µm). Cells were washed while still on the filter and then resuspended in 2 L of fresh medium

lacking α-factor at 30°C for 30 min or 45 min. Release was verified by visualisation of cell morphology and flow cytometry (Figure S2J). Two biological replicates were performed for each condition, showing high correlation (Fig S7B).

UV-Crosslinking and cDNA analysis (CRAC)

The CRAC protocol used in this study is derived from Granneman et al. (2009) with some modifications described in (Challal et al., 2018; Colin et al., 2014).

Briefly, cells were crosslinked by UV exposure for 50 seconds using a W5 UV crosslinking unit (UVO3 Ltd) and harvested by centrifugation at 4°C. Cell pellets were washed once with ice-cold 1x PBS, weighted and resuspended in 2.4 mL/(g of cells) of TN150 buffer (50 mM Tris pH 7.8, 150 mM NaCl, 0.1% NP-40 and 5 mM β-mercaptoethanol) supplemented with fresh protease inhibitors (AEBSF, Complete™ EDTA-free Protease Inhibitor Cocktail, Roche). Emulsions were snap-frozen in droplets in liquid nitrogen and cells subjected to cryogenic grinding using a Ball Mill MM 400 (5 cycles of 3 minutes at 20 Hz). The resulting frozen lysates were thawed on ice, treated with DNase I (165 units per gram of cells) incubated at 25°C for 1h to solubilize the chromatin and then clarified by centrifugation at 16 krpm for 30 min at 4°C.

RNA-protein complexes were affinity-purified with M-280 tosylactivated dynabeads coupled with rabbit IgGs (10 mg of beads per sample), washed with TN1000 buffer (50 mM Tris pH 7.8, 1 M NaCl, 0.1% NP-40 and 5 mM β-mercaptoethanol), and eluted by TEV protease digestion. RNAs were subjected to partial degradation by treating with 0.2 U of Rnase cocktail (Rnace-IT, Agilent) and the reaction was stopped by the addition of guanidine–HCl to a final concentration of 6 M. Eluates underwent then a second immobilisation on Ni-NTA columns (Qiagen, 100 µl of slurry per sample) overnight at 4°C and were extensively washed. Sequencing adaptors were ligated to the RNA molecules as described in the original procedure. RNA-protein complexes were eluted with elution buffer containing 50 mM Tris pH 7.8, 50 mM NaCl, 150 mM imidazole, 0.1% NP-40 and 5 mM β-mercaptoethanol fractionated

using a Gel Elution Liquid Fraction Entrapment Electrophoresis (GelFree) system (Expedeon) following manufacturer's specifications. The fractions containing the protein of interest were treated with 100 µg of proteinase K, and RNAs were purified and reverse-transcribed using reverse transcriptase Superscript IV (Invitrogen).

After quantification of the recovered material via quantitative PCR, the cDNAs were amplified with an appropriate number of PCR cycles using LA Taq polymerase (Takara), and then the reactions were treated with 200 U/mL of Exonuclease I (NEB) for 1 h at 37°C. Finally, the DNA was purified using NucleoSpin columns (Macherey-Nagel) and sequenced on a NextSeq 500 Illumina sequencer.

The H-CRAC protocol contained a few modifications to improve the recovery of tagged RNases H. DNase I treatment was replaced by a step of chromatin shredding by sonication in an ice-cold bath (15 min, High, 45 sec ON/OFF, Diagenode). The GelFree fractionation was omitted to avoid loss of material because the eluate after the second purification step was judged sufficiently pure.

Copy Number analysis

An aliquot roughly corresponding to 0.2 g of lysate powder from the CRAC experiment was transferred to a separate tube and used to perform genomic extraction using the Genomic-tip 20/G kit (QIAGEN) following manufacturer's specifications. DNA was fragmented using sonication (~200 to 500 bp size range). Sequencing libraries were prepared using a ThruPLEX DNA-seq kit (Rubicon Genomics). Next-generation sequencing was performed on a HiSeq 4000 (Illumina). Single-end reads of 50 bp were aligned to the *S. cerevisiae* genome (2011).

DNA:RNA immunoprecipitation (DRIP)

DNA:RNA hybrid immunoprecipitation (DRIP) was performed using the S9.6 DNA:RNA hybrid-specific monoclonal antibody according to a published procedure (Mischo et al., 2011; Wahba et al., 2016), with the modifications described in (Bonnet et al., 2017). Briefly, genomic

DNA was phenol-extracted from cells growing exponentially and isolated by ethanol precipitation. 50 µg of purified nucleic acids were digested by a cocktail of restriction enzymes (EcoRI, HindIII, XbaI, SspI, BsrGI; FastDigest enzymes; Thermo Scientific) for 30min at 37°C in a total volume of 100 µL. An RNase H treatment (10 units, Sigma) was included in the restriction reaction of control samples to assess the specificity of the DRIP signal. Digested samples were further diluted 4-fold with FA1 buffer (0.1% SDS, 1% Triton, 10 mM HEPES pH 7.5, 0.1% sodium deoxycholate, 275 mM NaCl) and incubated overnight at 4°C in the presence of 1.5 µg of S9.6 purified antibody (Kerafast). Antibody-associated DNA:RNA hybrids were then captured on protein G Sepharose beads (GE Healthcare), washed and purified according to standard ChIP procedures. Input and immunoprecipitated DNA amounts were quantified by real-time PCR with a LightCycler 480 system (Roche) using SYBR Green incorporation according to the manufacturer's instructions. Oligos DL4519 and DL4520 were used to amplify the mAIRN *locus* while oligos DL4597 and DL4598 were used to amplify an intergenic region located in proximity to the *HO* gene and used a negative control. The amount of DNA in the immunoprecipitated fraction was divided by the amount detected in the input to evaluate the percentage of immunoprecipitation (% of IP).

Imaging of Rad52-YFP *foci*

Rad52-YFP *foci* formation was assessed in exponentially growing cells ($0.5 \leq OD_{600} \leq 1$) in CSM medium at 30°C or 37°C as indicated. For wild-type Sen1-depleted conditions, Indole-3'-Acetic Acid (IAA, Sigma) was supplemented at a final concentration of 500 µM 1 h before imaging. Wide-field fluorescence images were acquired using a Leica DM6000B microscope with a 100X/1.4 NA (HCX Plan-Apo) oil immersion objective and a CCD camera (CoolSNAP HQ; Photometrics). The acquisition system was piloted by the MetaMorph software (Molecular Devices). For all images, z stacks sections of 0.2 µm were acquired using a piezo-electric motor (LVDT; Physik Instrument) mounted underneath the objective lens. Images were scaled equivalently and z-projected using ImageJ. An average of three experiments, each of them visualizing at least 300 cells per condition, is shown in the figures.

For thermosensitive alleles, and their relative controls, after reaching exponential growth cells were shifted to 37°C by the addition of pre-warmed media, and incubated for 1 h before imaging.

Protein analyses

Proteins levels were analysed using current methodologies.

Dataset processing and data analysis

CRAC

CRAC datasets were analysed as described (Candelli et al., 2018; Challal et al., 2018). The pyCRAC script pyFastqDuplicateRemover was used to collapse PCR duplicates using a 6 nucleotides random tag included in the 3' adaptor (see Key Resources Table). The resulting sequences were reverse complemented with Fastx reverse complement (part of the fastx toolkit, http://hannonlab.cshl.edu/ fastx_toolkit/) and mapped to the R64 genome (Cherry et al., 2012) with bowtie2 (-N 1) (Langmead and Salzberg, 2012).

Quantification and statistical analysis

The vast majority of the analyses were performed with inhouse scripts in the R studio environment. Sen1-AID CRAC datasets were analysed using the Galaxy web platform at usegalaxy.org (Afgan et al., 2018).

For all RNAPII CRAC data, the working group of 3000 genes with the highest expression level was selected by computing HT-seq count normalised to the size of the gene. This allowed excluding from our analysis genes with very low or background signal, which are potential source of computational biases.

The skew index was defined as the ratio between the RNAPII CRAC signals in the windows [0; +200]/[+300; +500] relative to the TSS for each gene. A *sen1-3*/WT skew index ratio was used to select genes with increased 5'-end RNAPII accumulation in the mutant

relative to the wild-type. All features with a skew index ratio exceeding the mean plus one standard deviation of the distribution were considered to be affected.

For Copy Number analysis, all regions with a score >1 were considered as undergoing replication. For the selection of the “already replicated genes” genes overlapping with a replicated region with a copy number score exceeding the 95th percentile were chosen.

When average values were represented, error bars indicate standard deviation. T tests were used to compare distributions and p-values are indicated.

KEY RESOURCE TABLE

Antibodies		
Reagent or Resource	**Source**	**Identifier**
IgG from rabbit serum	Sigma-Aldrich	Cat# I5006; RRID: AB_1163659
Mouse anti Flag	Sigma-Aldrich	Cat# F1804; RRID: AB_262044
Rabbit Peroxidase Anti-Peroxidase	Sigma-Aldrich	Cat# P1291; RRID: AB_1079562
Anti-phospho Histone H2A (Ser129)	Merck	Cat# 07-745-I
Anti-DNA:RNA hybrids (S9.6)	Kerafast	Cat# ENH001; RRID: AB_2687463
Goat anti-rabbit IgG-HRP	Santa Cruz	Cat# sc-2004; RRID: AB_631746

Chemicals, Peptides, and Recombinant Proteins		
Reagent or Resource	**Source**	**Identifier**
cOmplete EDTA-free protease inhibitor cocktail tablets	Sigma-Aldrich (Roche)	Cat# 11873580001
Pefabloc SC-Protease-Inhibitor	Carl Roth	Cat# A154.3
Dnase I recombinant, Rnase-free	Sigma-Aldrich (Roche)	Cat# 04716728001

Dynabeads M-280 Tosylactivated	Thermo Fisher Scientific	Cat# 14204
Recombinant GST-TEV protease	(Challal et al., 2018)	N/A
Rnace-It Ribonuclease Cocktail	Agilent	Cat# 400720
Guanidine hydrochloride	Sigma-Aldrich	Cat# G4505
Ni-NTA Agarose	QIAGEN	Cat# 30230
Imidazole	Sigma-Aldrich	Cat# I0125
RNaseOUT Recombinant Ribonuclease Inhibitor	Thermo Fisher Scientific	Cat# 10777019
T4 RNA Ligase 2, truncated KQ	NEB	Cat# M0373L
T4 Polynucleotide Kinase	NEB	Cat# M0201L
T4 RNA Ligase 1 (ssRNA Ligase)	NEB	Cat# M0204L
Proteinase K, recombinant, PCR grade	Sigma-Aldrich (Roche)	Cat# 03115887001
SuperScript IV Reverse Transcriptase	Thermo Fisher Scientific	Cat# 18090050
Rnase H	NEB	Cat# M0297S
Exonuclease I	NEB	Cat# M0293S
LA Taq	Takara	Cat# RR002M
FastDigest EcoRI	Thermo Fisher Scientific	Cat# FD0275
FastDigest HindIII	Thermo Fisher Scientific	Cat# FD0505
FastDigest XbaI	Thermo Fisher Scientific	Cat# FD0685
FastDigest SspI	Thermo Fisher Scientific	Cat# FD0774
FastDigest Bsp1407I (BsrGI)	Thermo Fisher Scientific	Cat# FD0934
Rnase H	Sigma-Aldrich (Roche)	Cat# 10786357001
Protein G Sepharose Fast Flow	GE Healthcare	Cat# 17061801
Hydroxyurea (HU)	Sigma-Aldrich	Cat# H8627
Methyl methanesulfonate (MMS)	Sigma-Aldrich	Cat# 129925
α-factor	BIOTEM	N/A
Paraformaldehyde	VWR Chemicals	Cat# 28794.295
3-Indoleacetic acid (IAA)	Sigma-Aldrich	Cat# I2886

Strains			
Lab number	**Identifier**	**Source**	**Genotype**
DLY671	Wild-type	F. Lacroute	*as BMA64; Mat a*
DLY128	*rna15-2*	F. Lacroute	*as W303; rna15-2; Mat a*
DLY753	*rat1-1*	F. Lacroute	*as BMA64 ; rat1-1 ; Mat alpha*
DLY2057	*sen1-1*	F. Lacroute	*as BMA64; sen1-1; Mat a*
DLY2571	Rpb1-HTP	(Candelli et al., 2018)	*as BMA64; RPB1::HTP::TRP1kl; Mat a*
DLY3173	*sen1-3*-TAP	(Appanah et al., 2020)	*as W303; sen1 W773A E774A W777A::TAP::KanMX; Mat a*
DLY3197	*sen1-3*	This study	*as W303; sen1 W773A E774A W777A; Mat a*
DLY3211	Rpb1-HTP *sen1-3*	This study	*as BMA64; sen1 W773A E774A W777A; RPB1:HTP::TRP1kl; Mat a*
DLY3321	Rnh1-HTP	This study	*as BMA64; RNH1::HTP::TRP1kl; Mat a*
DLY3348	Rnh1-HTP *sen1-3*	This study	*as BMA64; RNH1::HTP::TRP1kl; sen1 W773A E774A W777A; Mat a*
DLY3368	*sen1-3*-TAP *rat1-1*	This study	*as BMA64; sen1 W773A E774A W777A::TAP::KanMX; rat1-1; Mat a*
DLY3370	*sen1-3*-TAP *rat15-2*	This study	*as BMA64; sen1 W773A E774A W777A::TAP::KanMX; rna15-2; Mat a*
DLY3421	Rpb1-HTP *rnh1Δ rnh201Δ*	This study	*as BMA64; rnh1::hphNT; rnh20http:isMX; RPB1::HTP::TRP1kl; Mat a*
DLY3432	Rnh201-HTP	This study	as *BMA64; RNH201::HTP::TRP1kl; Mat a*
DLY3438	Rnh201-HTP *sen1-3*	This study	*as BMA64; RNH201::HTP::TRP1kl; sen1 W773A E774A W777A; Mat a*
DLY3443	Rad52-YFP	(Lisby et al., 2001)	*as BMA64; RAD52::YFP; bar1::LEU2; Mat a*

DLY3477	Rad52-YFP Sen1-AID pFL38-*sen1-3*	This study	*as BMA64; SEN1::AID::KAN::OsTIR1; RAD52::YFP; pFL38-sen1-3::URA; Mat a*
DLY3479	Rad52-YFP *rnh1Δ rnh201Δ*	This study	*as BMA64; rnh1::hphNT; rnh20http:isMX; RAD52::YFP; RPB1::HTP::TRPkl; bar1::LEU2; Mat a*
DLY3481	Rad52-YFP Sen1-AID *rnh1Δ rnh201Δ* + pFL38-*sen1-3*	This study	*as BMA64; SEN1::AID::KAN::OsTIR1; rnh1::hphNT; rhttp:1::HisMX; RAD52::YFP; RPB1::HTP::TRPkl; bar1::LEU2; pFL38-sen1-3::URA; Mat a*
DLY3562	Rad52-YFP *sen1-3*	This study	*As W303, sen1-3, RAD52::YFP*
DLY3582	Rad52-YFP *sen1-1*	This study	*As W303, sen1-1, RAD52::YFP*
DLY3583	Rad52-YFP *nrd1-102*	This study	*As W303, nrd1-102, RAD52::YFP*
DLY3584	Rad52-YFP *sen1-3 nrd1-102*	This study	*As W303, nrd1-102, sen1-3, RAD52::YFP*

Critical commercial assay		
Reagent or Resource	**Source**	**Identifier**
Genomic-tip 20/G kit	QUIAGEN	Cat# 10223
LightCycler 480 SYBR Green I Master	Roche	Cat# 04887352001
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel	Cat# 740609
Pierce Spin Columns - Snap Cap	Thermo Fisher Scientific	Cat# 69725
Vivacon 500	Sartorius	Cat# VN01H22

Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific (Invitrogen)	Cat# Q32851
SuperSignal West Pico Chemiluminescent Substrate	Thermo Fisher Scientific	Cat# 34080

Plasmids		
Reagent or Resource	**Source**	**Identifier**
pRS314-L	(Prado and Aguilera, 2005)	N/A
pDL983-pCM190-mAIRN-350to846-pGAL-CUP1	This study	N/A
pDL987-pRS424-GPDprom-hs—H1	(Wahba et al, 2011)	N/A
pDL1010-pRS424-GPDprom-PTH-hsRNH1	This study	N/A

Software and Algorithms		
Reagent or Resource	**Source**	**Identifier**
RStudio	RStudio	RRID:SCR_000432
ImageJ	NIH	https://imagej.nih.gov/ij/; RRID:SCR_003070
Affinity Designer	Serif	https://affinity.serif.com/en-us/designer/; RRID:SCR_016952

Oligonucleotides		
Reagent or Resource	**Source**	**Identifier**

DL4519- GGTTTACGGGCGATTTAGAGCA	This study	N/A
DL4520- CAACTCTCCAGCAGCGTGGT	This study	N/A
DL4597- GAAACCACGAAAAGTTCACCA	This study	N/A
DL4598- AGCTTCTGCAAACCTCATTTG	This study	N/A

DATA AVAILABILITY

All the datasets generate by this study have been deposited in the Geo Expression Omnibus repository at NCBI and are available using the code GSE195936.

ACKNOWLEDGEMENTS

We wish to thank Andres Aguilera, Doug Koshland, Rodney Rothstein for providing strains or plasmids; Giacomo De Piccoli, Sarah Lambert, Vincent Vanoosthuyse, Michel Werner, Frédéric Chédin, Julien Gros and the members of the Libri and Palancade labs for critical reading of the manuscript and fruitful discussion; Vasudha Sharma for help with the generation of mutants and with microscopy experiments. We also wish to thank Yan Jaszczyszyn for expert technical help in preparing CRAC libraries for sequencing. This work has benefited from the facilities and expertise of the high throughput sequencing core facility of I2BC (Centre de Recherche de Gif - http://www.i2bc-saclay.fr/).

AUTHOR CONTRIBUTION

UA and DL designed experiments. UA performed all the experiments and analysis excluding the raw data analysis of the DNA Copy number and the Sen1-AID RNAPII CRAC and the ensuing bioinformatic analysis presented in Figure 1. DC performed the Sen1-AID RNAPII CRAC. GW offered technical assistance for several experiments. AL assisted UA with

the DNA Copy Number analysis. BP assisted UA with DRIP experiments and Rad52 *foci* imaging. RA provided the sen1-3 strain ahead of publication. DL performed the bioinformatic analysis presented in Figure 1. UA and DL wrote the manuscript. DL supervised the project. DL collected funding. All authors provided feedbacks during the writing and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

FUNDING

This work was supported by the Centre National de la Recherche Scientifique (C.N.R.S.), the Fondation pour la Recherche Medicale (F.R.M., programme Equipes 2019 to D.L.), l'Agence National pour la Recherche (ANR-16-CE12-0022-01 to D.L. and P.P.; ANR-21-CE12-0040-01 to D.L. and B.P.; ANR-18-CE12-0003 to B.P.; ANR-19-CE12-0016-01 to P.P.), the Fondation ARC pour la Recherche sur le Cancer and the Ligue Nationale contre le Cancer to B.P. and the IdEx Université de Paris (ANR-18-IDEX-0001 to B.P.).

U.A. was supported by the French Ministry for Education and Research, by the Fondation ARC pour la Recherche sur le Cancer and the EUR G.E.N.E. (reference #ANR-17-EURE-0013), which is part of the Université de Paris IdEx #ANR-18-IDEX-0001 funded by the French Government through its "Investments for the Future" program.

REFERENCES

Achar, Y.J., Adhil, M., Choudhary, R., Gilbert, N., and Foiani, M. (2020). Negative supercoil at gene boundaries modulates gene topology. Nature *577*, 701–705.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. *46*, W537–W544.

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

Appanah, R., Lones, E.C., Aiello, U., Libri, D., and De Piccoli, G. (2020). Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability. Cell Rep. *30*, 2094-2105.e9.

Bohnsack, M.T., Tollervey, D., and Granneman, S. (2012). Identification of RNA helicase target sites by UV cross-linking and analysis of cDNA. Methods Enzymol. *511*, 275–288.

Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Géli, V., de Almeida, S.F., and Palancade, B. (2017). Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability. Mol. Cell *67*, 608-621.e6.

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

Carrasco-Salas, Y., Malapert, A., Sulthana, S., Molcrette, B., Chazot-Franguiadakis, L., Bernard, P., Chédin, F., Faivre-Moskalenko, C., and Vanoosthuyse, V. (2019). The extruded non-template strand determines the architecture of R-loops. Nucleic Acids Res. *47*, 6783–6795.

Castel, S.E., Ren, J., Bhattacharjee, S., Chang, A.-Y., Sánchez, M., Valbuena, A., Antequera, F., and Martienssen, R.A. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. Cell *159*, 572–583.

Challal, D., Barucco, M., Kubik, S., Feuerbach, F., Candelli, T., Geoffroy, H., Benaksas, C., Shore, D., and Libri, D. (2018). General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. Mol. Cell *72*, 955-969.e7.

Chédin, F., Hartono, S.R., Sanz, L.A., and Vanoosthuyse, V. (2021). Best practices for the visualization, mapping, and manipulation of R-loops. EMBO J. *40*, e106394.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, D700-705.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

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

Costantino, L., and Koshland, D. (2018). Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. Mol. Cell *71*, 487-497.e3.

Delan-Forino, C., Schneider, C., and Tollervey, D. (2017). Transcriptome-wide analysis of alternative routes for RNA substrates into the exosome complex. PLoS Genet. *13*, e1006699.

Feng, Y., Seija, N., Di Noia, J.M., and Martin, A. (2020). AID in Antibody Diversification: There and Back Again. Trends Immunol. *41*, 586–600.

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

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc. Natl. Acad. Sci. *106*, 9613–9618.

Haeusler, R.A., and Engelke, D.R. (2006). Spatial organization of transcription by RNA polymerase III. Nucleic Acids Res. *34*, 4826–4836.

Han, Z., Libri, D., and Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. Nucleic Acids Res. *45*, 1355–1370.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. Mol. Cell *31*, 925–932.

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2012). Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. Mol Cell *49*, 55–66.

Kanagaraj, R., Mitter, R., Kantidakis, T., Edwards, M.M., Benitez, A., Chakravarty, P., Fu, B., Becherel, O., Yang, F., Lavin, M.F., et al. (2022). Integrated genome and transcriptome analyses reveal the mechanism of genome instability in ataxia with oculomotor apraxia 2. Proc. Natl. Acad. Sci. *119*.

Kawauchi, J., Mischo, H., Braglia, P., Rondon, A., and Proudfoot, N.J. (2008). Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. Genes Dev. *22*, 1082–1092.

Kobayashi, T. (2003). The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. Mol. Cell. Biol. *23*, 9178–9188.

Kobayashi, T., and Ganley, A.R.D. (2005). Recombination regulation by transcription-induced cohesin dissociation in rDNA repeats. Science *309*, 1581–1584.

Lai, F., Damle, S.S., Ling, K.K., and Rigo, F. (2020). Directed RNase H Cleavage of Nascent Transcripts Causes Transcription Termination. Mol. Cell.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Leonaitė, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O., and Conti, E. (2017). Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. EMBO J. *36*, 1590–1604.

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

Longtine, M.I., Demarini, S., Wach, B., and Philippsen, P. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. 9.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2011). Uniform transitions of the general RNA polymerase II transcription complex. Nat Struct Mol Biol *17*, 1272–1278.

Mendoza-Ochoa, G.I., Barrass, J.D., Terlouw, B.R., Maudlin, I.E., de Lucas, S., Sani, E., Aslanzadeh, V., Reid, J.A.E., and Beggs, J.D. (2019). A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. Yeast Chichester Engl. *36*, 75–81.

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

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

Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nat. Struct. Mol. Biol. *20*, 884–891.

Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends Genet. TIG *32*, 508–522.

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

Rossi, M.J., Kuntala, P.K., Lai, W.K.M., Yamada, N., Badjatia, N., Mittal, C., Kuzu, G., Bocklund, K., Farrell, N.P., Blanda, T.R., et al. (2021). A high-resolution protein architecture of the budding yeast genome. Nature *592*, 309–314.

San Martin-Alonso, M., Soler-Oliva, M.E., García-Rubio, M., García-Muse, T., and Aguilera, A. (2021). Harmful R-loops are prevented via different cell cycle-specific mechanisms. Nat. Commun. *12*, 4451.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-wide mapping of yeast RNA polymerase II termination. PLoS Genet. *10*, e1004632.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075–1087.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol Cell *24*, 735–746.

Thompson, M., Haeusler, R.A., Good, P.D., and Engelke, D.R. (2003). Nucleolar Clustering of Dispersed tRNA Genes. Science *302*, 1399–1401.

Tye, B.W., Commins, N., Ryazanova, L.V., Wühr, M., Springer, M., Pincus, D., and Churchman, L.S. (2019). Proteotoxicity from aberrant ribosome biogenesis compromises cell fitness. ELife *8*, e43002.

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

Wang, S., Han, Z., Libri, D., Porrua, O., and Strick, T.R. (2019). Single-molecule characterization of extrinsic transcription termination by Sen1 helicase. Nat. Commun. *10*, 1545.

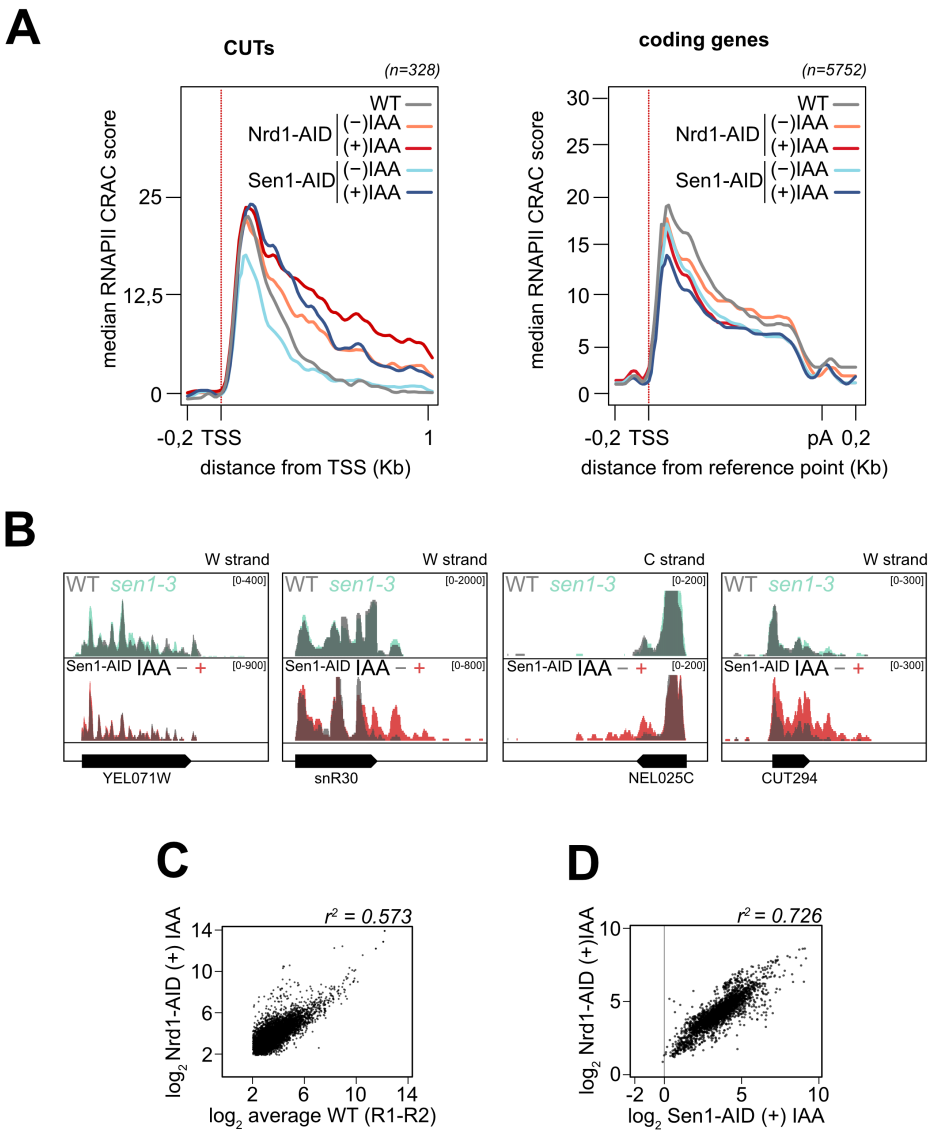
Xie, J., Aiello, U., Clement, Y., Haidara, N., Girbig, M., Schmitzova, J., Pena, V., Müller, C.W., Libri, D., and Porrua, O. (2021). An integrated model for termination of RNA polymerase III transcription.

Zardoni, L., Nardini, E., Brambati, A., Lucca, C., Choudhary, R., Loperfido, F., Sabbioneda, S., and Liberi, G. (2021). Elongating RNA polymerase II and RNA:DNA hybrids hinder fork progression and gene expression at sites of head-on replication-transcription collisions. Nucleic Acids Res. *49*, 12769–12784.

Zencir, S., Dilg, D., Rueda, M.P., Shore, D., and Albert, B. (2020). Mechanisms coordinating ribosomal protein gene transcription in response to stress. Nucleic Acids Res. *48*, 11408–11420.

SUPPLEMENTAL FIGURES

Figure S1 related to Figure 1

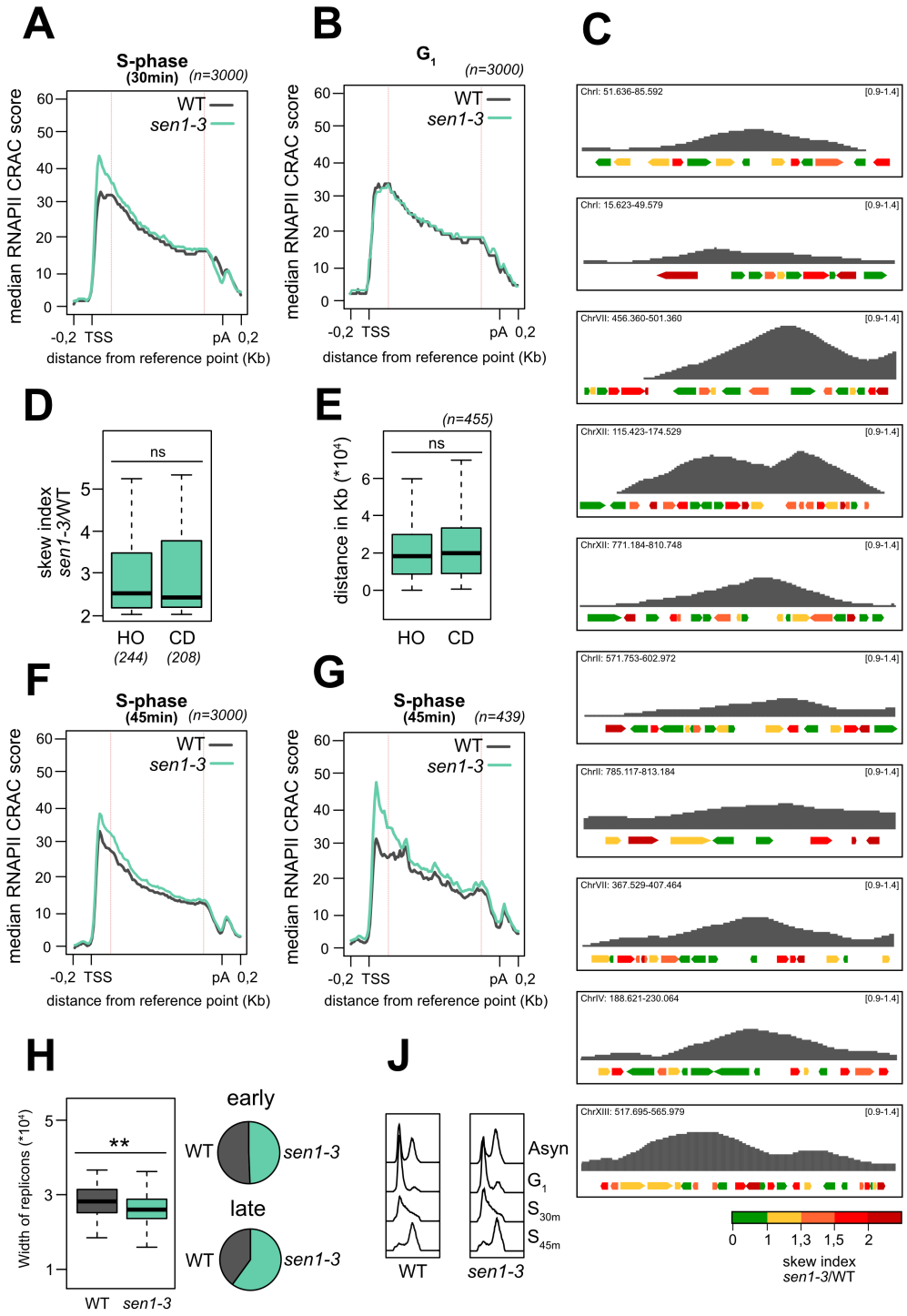


Supplementary Figure 1 related to Figure 1:

A) Metagene analysis of RNAPII distribution at a subset of validated Cryptic Unstable Transcripts (CUTs**,** left panel) and at coding genes (right panel) aligned at their Transcription Start Site (TSS) in wild-type (WT), Nrd1-AID and Sen1-AID cells in presence or absence of auxin (IAA). Values on the y-axis correspond to the median coverage. Note that the 3'-end of CUTs is not well defined, hence the increase in RNAPII occupancy upon depletion of Nrd1 or Sen1 is spread over a large region. **B)** Representative snapshots illustrating the absence of termination defects at coding and non-coding genes in *sen1-3* cells. For comparison, the

tracks derived from Sen1-AID cells in presence or absence of auxin (IAA), showing read-through at CUTs and snoRNAs are also included. *YEL071W* is shown as a representative example of lack of termination defects at coding genes in Sen1-AID cells even in presence of IAA. **C)** Scatter plot as in Figure 1A but for Nrd1-AID strain under depletion conditions (auxin added for 1 hour). **D)** Scatter plot analysis showing the good correlation of RNAPII CRAC density in mRNA coding genes in Sen1- and Nrd1-depleted cells.

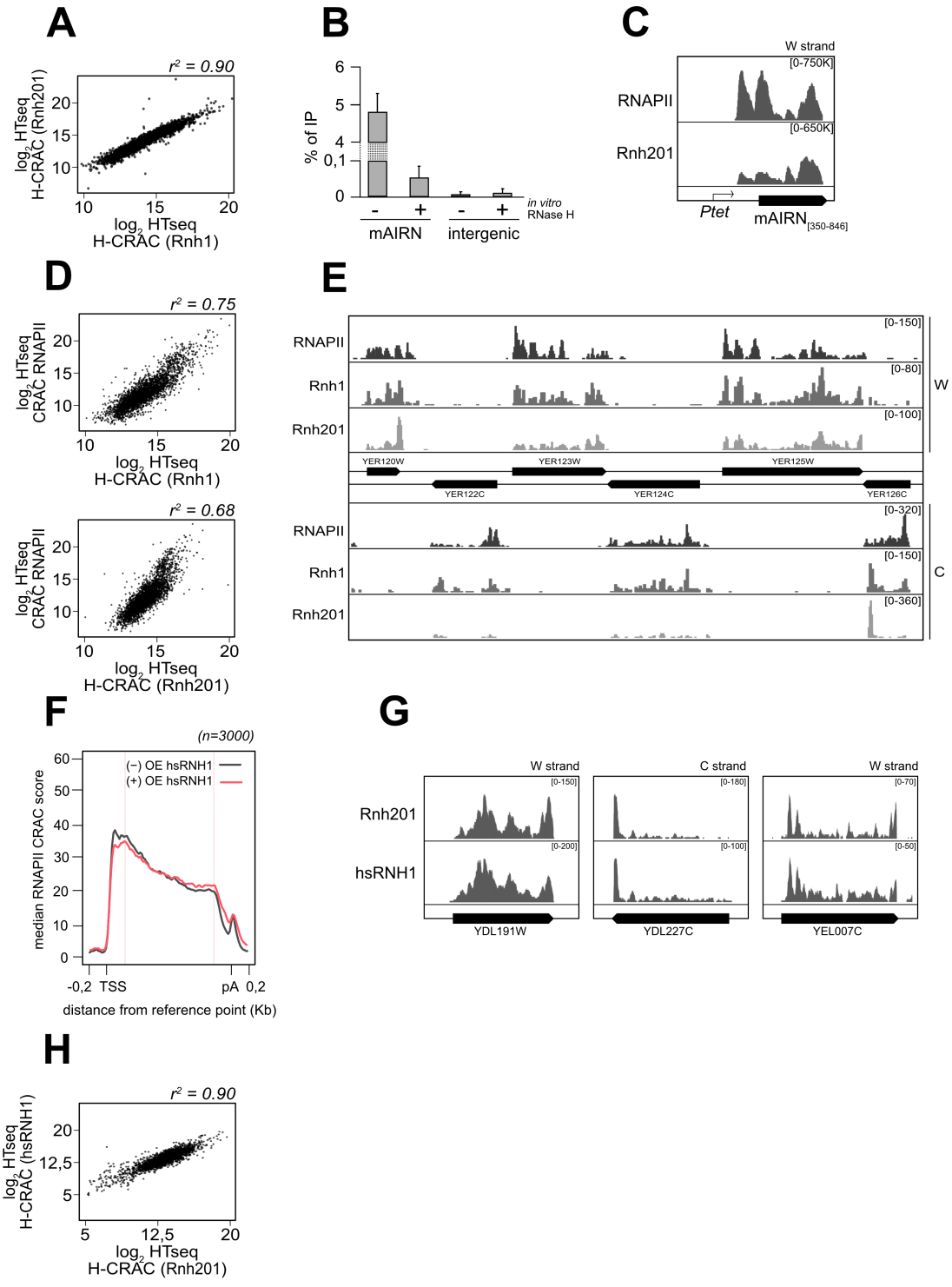
Figure S2 related to Figure 2



Supplementary Figure 2 related to Figure 2:

A) and **B)** As in Figure 2A, but on cells synchronously released in S-phase and collected 30 min after replication onset or arrested in G_1, respectively. The 3000 genes with the highest transcription levels have been used. **C)** As in Figure 2F, additional representative snapshots illustrating the frequent co-localisation of the replication forks (minima of the DNA copy number signal) and genes displaying a 5'-end skewed RNAPII distribution pattern in *sen1-3* cells (coloured in red). **D)** Boxplot comparison of the skew index ratio (*sen1-3*/WT*)* of RNAPII signal for genes oriented head-on (HO) or in co-direction (CD) relative to the position on the nearest activated ARS as detected by DNA copy number analysis. Genes in both groups have a replication-dependent accumulation of RNAPII in their 5'-end. **E)** Distribution of distances for each affected gene from the closest HO and CD origins. Affected genes are not preferentially replicated in one configuration over the other. **F)** Meta-analyses as in Figure 2A, but on cells synchronously released in S-phase and collected 45 min after replication onset. **G)** As in Figure 2C, but on cells synchronously released in S-phase and collected 45 min after replication onset. The analysis was performed on the subset of genes affected at this time point. **H)** Left: comparison of the width of the replicons detected by DNA copy number analysis from wild-type (WT) and *sen1-3* cells synchronously grown in S-phase and collected 30 min after replication onset. ** $p<0.01$. Right: pies indicating the total number of activated origins that were retrieved from DNA copy number analysis in wild-type (WT) and *sen1-3* cells. Origins were divided in early and late according to their replication timing in wild-type cells. **J)** Examples of cell cycle analysis by flow cytometry from the cells used for the experiments shown in Figure 2 and S2.

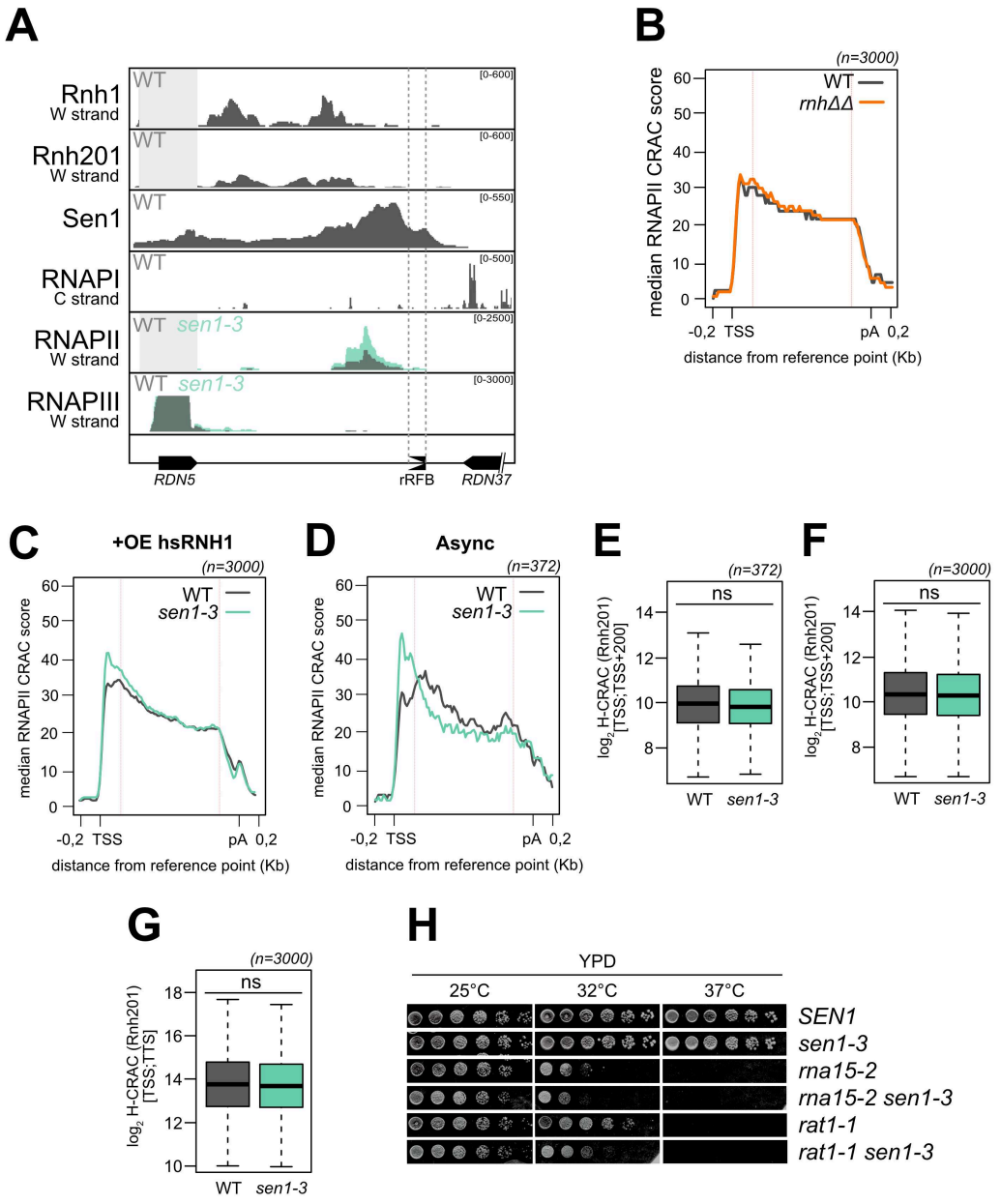
Figure S3 related to Figure 4



Supplementary Figure 3 related to Figure 4:

A) Dispersion plot of the log_2 values obtained from Rnh1 H-CRAC and Rnh201 H-CRAC for the 3000 most transcribed genes in asynchronous cells. The coefficient of determination (r^2) is shown. **B)** DNA:RNA Immunoprecipitation followed by quantitative PCR (DRIP-qPCR) from cells transformed with a plasmid carrying the mAIRN sequence expressed under control of the *pTet* promoter as indicated on the scheme in panel C (only the R-loop-forming region corresponding to the 350-848 nt interval of the mouse gene was cloned). The percentage of immunoprecipitated material is plotted on y-axis. The DRIP signal from an intergenic region located nearby to the *HO* locus was used as a negative control. Samples were treated or not with RNase H *in vitro* prior to immunoprecipitation as indicated. **C)** Read coverage for RNAPII (CRAC) and Rnh201 (H-CRAC) on the plasmid-borne mAIRN sequence. **D)** Dispersion plots illustrating the correlation between transcription in genes (RNAPII CRAC) and R-loop levels, as determined by H-CRAC (top: Rnh1; bottom: Rnh201). **E)** Integrative Genomics Viewer (IGV) representative screenshot of a chromosomic region illustrating the marked directionality of H-CRAC signals for both Rnh1 and Rnh201 as indicated. **F)** Metagene analyses of RNAPII CRAC signal at coding genes aligned on their TSS and on their pA site in wild-type (WT) cells transformed with a plasmid overexpressing hsRNH1 (+) or an empty plasmid (-). Genes are only scaled in between the red lines. **G)** Representative snapshots illustrating the similarities between the H-CRAC signal obtained with yeast Rnh201 or human RNH1 (hsRNH1). **H)** Dispersion plot as in Figure S3A but comparing Rnh201 to hsRNH1.

Figure S4 related to Figure 4

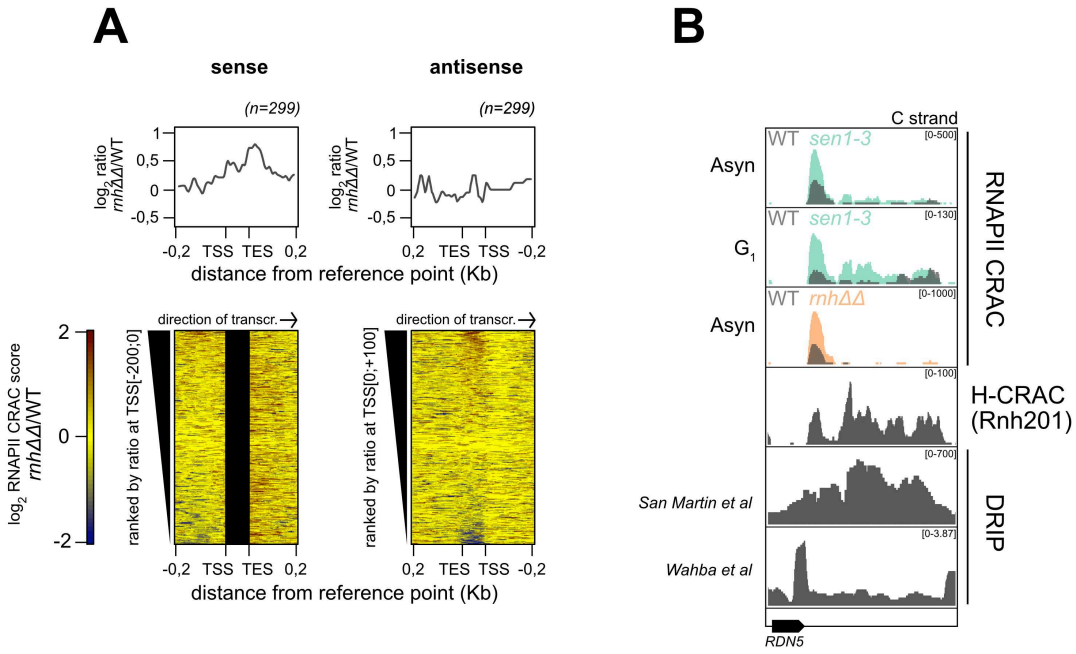


Supplementary Figure 4 related to Figure 4:

A) Distribution of CRAC signals with the indicated proteins at the rRFB. Data for RNAPI (Turowski et al., 2020) and RNAPIII (Xie et al., 2021) indicate that accumulation of RNAPII is unlikely due to conflicts with RNAPI or RNAPIII. Sen1 ChIP-exo signals (Rossi et al., 2021) demonstrate the presence of Sen1 at the site of TRC. The strand shown is indicated for each protein, with the exception of Sen1 because ChIP data are not directional. **B)** Metagene analyses of RNAPII CRAC signal at coding genes aligned on their TSS and pA site in *rnh1Δ*

rnh201Δ (*rnhΔΔ*) or wild-type (WT) cells. Genes are scaled only in between the red lines. **C)** As in Figure S4B but for WT and *sen1-3* cells transformed with a plasmid over-expressing human RNH1. Note that the 5'-end RNAPII accumulation observed in *sen1-3* cells is not lost in these conditions. **D)** As in Figure 2C but on asynchronously dividing cells and on the subset of genes affected in this condition. **E)** and **F)** Boxplot comparisons of the H-CRAC signal on the interval [TSS; TSS+200] for the same group of genes shown in Figure S4D or for the most transcribed 3000 genes, respectively. **G)** As in Figure S4F, but for the signal along the full gene [TSS; TTS]. **H)** Growth assay of *sen1-3 rat1-1* cells compared to single mutants. Serial dilutions of the indicated strains were incubated for 3 days at the indicated temperature. Growth was performed on the same plates.

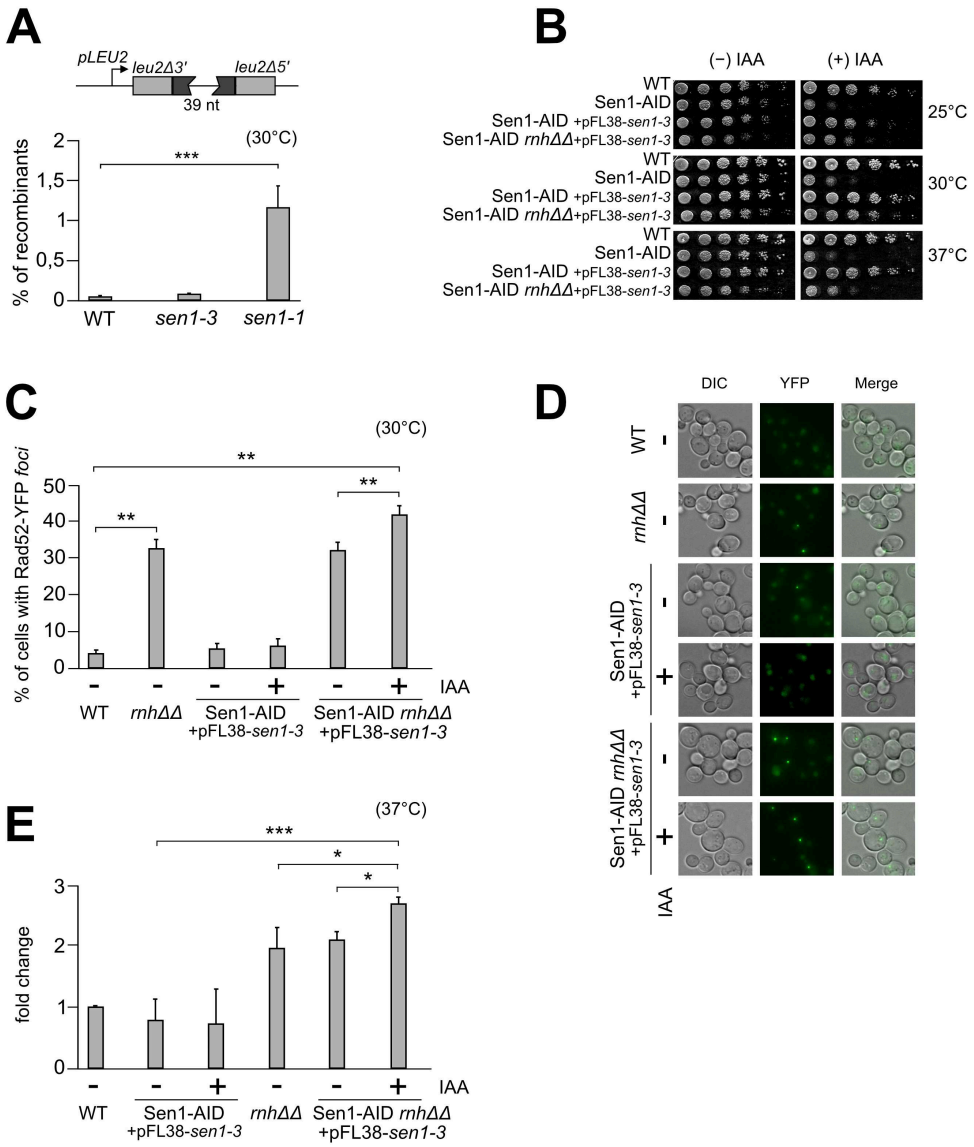
Figure S5 related to Figure 5



Supplementary Figure 5 related to Figure 5:

A) Heatmap analyses representing the $\log_2$ ratio (*rnh1Δ rnh201Δ* /WT*)* of the RNAPII CRAC signal at tRNA genes aligned at their Transcription Start Site (TSS) both for the sense and the antisense transcription as indicated. Genes were ranked as in Figure 5B. The summary plot on the top was calculated using the median values foreach position. **B)** Accumulation of RNAPII antisense to the *RDN5* gene in WT, *sen1-3* and *rnh1Δ rnh201Δ* cells as in Figure 5A for tRNAs. R-loops levels from H-CRAC and DRIP-seq (San Martin-Alonso et al., 2021; Wahba et al., 2016) are also shown.

Figure S6 related to Figure 6

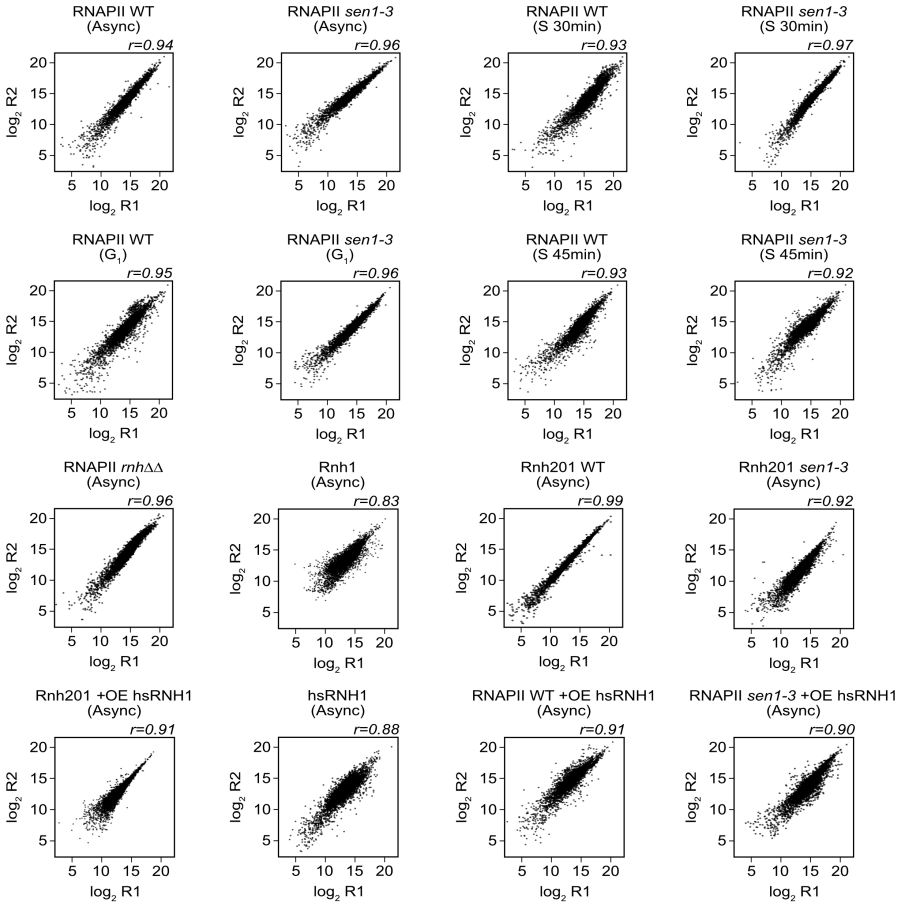


Supplementary Figure 6 related to Figure 6:

A) Frequency of transcription-associated recombination (TAR) events assessed using the pL (Prado and Aguilera, 2005) reporter in wild-type (WT), *sen1-3* and *sen1-1* cells. The reporter plasmid (schematised above the graph) contains the *LEU2* gene interrupted by a 39 nt insertion flanked by homology repeats. Transcription activation in the absence of leucine induces damage and recombination between the two repeats, which reconstitutes a functional

LEU2 gene. *** p < 0.001 **B)** Growth assay of inducible triple mutant used for the analyses shown in Figures 6 and S6. Note that the establishment of the phenotype is partial, possibly due to the partial depletion of Sen1 or a suppression effect of slight Sen1-3 overexpression. **C)** As in Figure 6B, but growth was performed at 30°C. ** p< 0.01. **D)** Representative examples of the microscopy data shown in Figure S6C. **E)** Quantification of H2A Ser129 phosphorylation detected by western blot in Figure 6C. Cells were grown in logarithmic phase at 30°C, and then shifted at 37°C for 1 hour. The fold change relative to the WT levels is shown for the indicated conditions. Pab1, Nrd1 and Nab3 were used as loading controls. Error bars represent standard deviations. * p<0.05; *** p<0.001.

Figure S7



Supplementary Figure 7:

Correlation plots between replicates of the CRAC experiments shown in this study. The Pearson correlation score (r) is shown at the top right of each plot.

III. The 3'-end of genes is a main determinant of RNases H and Sen1 binding.

We have established that H-CRAC detects with high sensitivity and resolution RNase H targets, which at least partially coincide with R-loops. Many controls have been included in the characterisation of the method, as described in Results, Chapter II. The overall genome-wide analysis of R-loops/RNases H targets have not been described in the manuscript presented in Chapter II and they will be included in another study that will be submitted shortly. I will briefly describe here the main results of these analysis that is still ongoing. Because additional experiments and bioinformatic analyses are required, many of the conclusions reported below must be considered preliminary and might not be supported by future analyses.

All experiments and analysis presented in this section were performed by me.

Rnh1 and Rnh201 targets are located along genes and close to their pA site.

Several studies have shown that the double deletion of RNase H genes is required to trigger phenotypes such as R-loops accumulation and genetic instability, suggesting that the functions of RNase H1 and H2 are a least partially redundant. However, the two enzymes have also been proposed to play different roles in physiological conditions. In order to better understand the function of RNases H enzyme we set up to investigate their binding sites and specificities through the comparative analysis of our H-CRAC maps.

H-CRAC signal overlaps with transcription units

R-loops are by-products of RNAP activity and therefore it can be envisioned that RNases H binding should be dictated by the occurrence of transcription. Consistently, H-CRAC signal largely overlapped transcription units, with virtually no signal found in intergenic non-transcribed regions (Figure 22A). Importantly, signal could be detected

also in unstable transcripts such as CUTs or SUTs which are rapidly degraded after transcription termination and in introns, which are removed from the mature transcript (data not shown). This observation suggests that H-CRAC can reveal co-transcriptional R-loops. Nevertheless, the distribution of the genic signal for RNases H and RNAPII was different (Figure 22A), indicating that not all sequences form R-loops with equal efficiency.

H-CRAC signal is biased towards the 3'-end of coding genes

To better understand the targets of RNase H activity we first focused on coding genes. The H-CRAC methodology benefits from an unprecedented resolution, which allowed us to accurately detect the sites of RNase H binding, presumably R-loops. By inspecting our H-CRAC maps we noticed that many genes contain a prominent and well-defined peak, located close to the polyadenylation (pA) site. Importantly, these peaks are not generally found in the pattern observed for RNAPII CRAC, suggesting that they are strong sites of RNase H binding (Figure 22A). This observation prompted us to investigate more extensively the distribution of the H-CRAC signal along mRNA genes. Therefore, we compared the binding of RNases H at the 5'-end [TSS; TSS+150], in the gene body [TSS+150; pA-150] and at the 3'-end of genes [pA-150; pA] by measuring the density of signal (i.e., total read count divided by the size of the region) in these regions for all the coding genes. The 3'-end appeared to be the most prominent site, accounting for about half of the total signal in coding genes (Figure 22B). Such a biased distribution could be explained both by the presence of a group of genes with very strong signal at their 3'-end, or by a general tendency of all genes to display more signal at the 3'-end compared to the 5'-end or the body regions. To distinguish between these two hypotheses, we performed heatmaps ranking the genes according to the signal at the pA site. As observed in Figure 22C, a group of genes contains a strong signal at the pA site for both Rnh1 and Rnh201. However, this is not a generalised feature, and many other genes showed a more distributed H-CRAC signal without a 3'-end peak. Consistently, ranking the genes according to their signal all over the transcription unit did not recapitulate the same pattern observed by ranking for the signal at the 3'-end (Figure 22C).

H-CRAC defines two classes of genes: one with spread signal and one with a peak at their 3'-end

The same group of genes display a 3'-end peak of signal of both Rnh1 and Rnh201

Thus, we sought to identify the genes decorated with a peak in their 3' region. To this purpose we quantified for each gene the extent of the signal in the 3'-end region relative to the signal in the rest of the gene. This approach allowed us to select 911 and 795 genes, for Rnh1 and Rnh201 respectively, with the strongest 3'-end skewed distribution. Interestingly, intersection of these genes revealed a very significant overlap (570 genes, p-value = 3.05545e-319) (Figure 22D), showing that the vast majority of genes with a peak at their 3'-end are bound both by Rnh1 and Rnh201. Similarly, very strong overlap was found when intersecting genes selected for the total signal in the unit (1000 genes for each dataset) (Figure 22D). However, when intersecting the group of genes recovered by the two strategies (i.e., by selecting them for their total signal or for the presence of a peak in their 3'-end) for each dataset we noticed that the overlap was much lower (Figure 22D). These observations are in agreement with what we observed by heatmaps analyses (Figure 22C), and they overall suggest the existence of two different classes of genes in the dataset: a class with a signal spread along the transcription unit, and a class for which RNases H bind more specifically at the pA site, the two classes overlapping only to a limited extent.

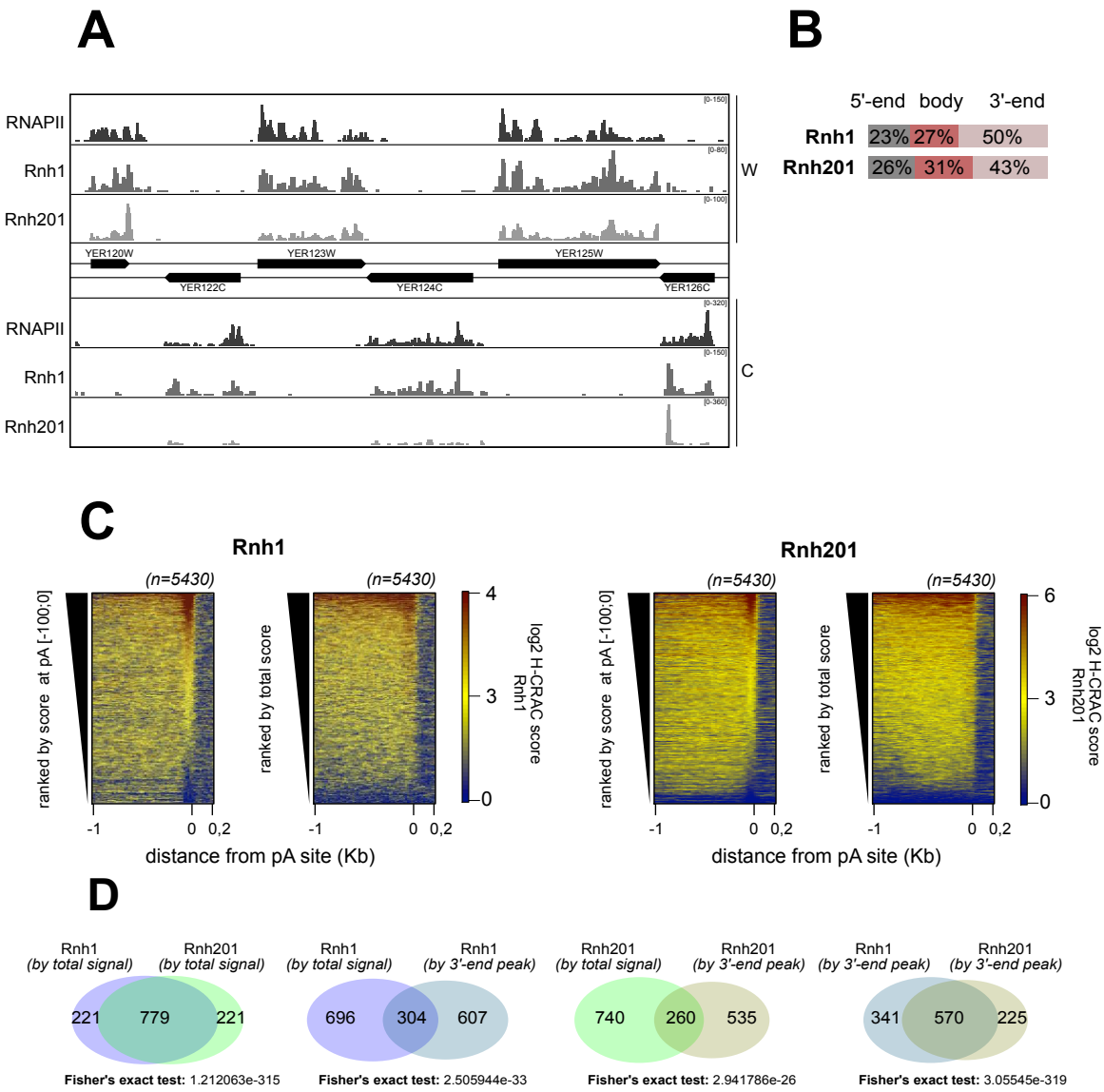


Figure 22. Rnh1 and Rnh201 targets are located along genes and close to their pA site.

A) Integrative Genomics Viewer (IGV) screenshot of a section of Chromosome V showing examples of H-CRAC and RNAPII signal at coding genes. The values in brackets correspond to the scale of the signal. The two strands are showed (W for Watson, C for Crick). The signal of a wild-type is shown.

B) Enrichment analysis of the H-CRAC signal for Rnh1 and Rhn201 along mRNA genes. The showed values are the normalised percentages of enrichment for each region.

C) Heatmap analysis representing the $\log_2$ the H-CRAC signal in coding genes aligned at their pA in wild-type cells. Genes were ranked following the order indicated on the left of each panel.

D) Overlap between the groups of genes enriched for a spread signal or for a peak at their 3'-end. The p-values calculated by Fisher's exact test are shown for each overlap.

Transcription, nucleotide composition and topological stress likely impact RNases H binding.

To better characterise the class of genes with a 3'-end skewed signal and address the mechanisms underlying the formation of the peak, we performed a metagene analyses on genes aligned on the pA site (Figure 23A). This allowed appreciating the average position of the peak and its characteristics, including average level and width which can be estimated about 100 nt, as also suggested by previous heatmap analysis (Figure 22C).

Genes with a H-CRAC 3'-end peak are more enriched in CG nucleotides

Nucleotide composition has been identified as one of the features influencing R-loop formation, being G-rich regions favoured (Ginno et al., 2013) (see Introduction, § III.I). Consequently, we calculated the nucleotide frequencies in the 100 nt region centred at the highest value for each peak and compared it to the pA region of other genes lacking an H-CRAC 3'-peak signal. A shift in nucleotide frequencies was observed (Figure 23B), with a statistically significant increase in the frequency of G and C nucleotides, at the expenses of A nucleotides. On the contrary, no difference in sequence composition was observed for the genes enriched with a signal all over the transcription unit, for which the nucleotide frequencies did not differ from those observed at genes with low H-CRAC signal (data not shown). We concluded that a GC sequence bias is likely influencing the formation of R-loops at the pA site.

Transcription rate explains a good portion of the variability observed in H-CRAC signal

Transcription rate is another major determinant of R-loop levels; therefore, it is expected that H-CRAC signals have strong correlation with transcription levels. Using RNAPII CRAC density as a mean to estimate the transcription rate, we generated scatter plots using the RNAPII CRAC and the relative H-CRAC score for both Rnh1 and Rhn201. The points were then fitted on a linear model and Pearson correlation scores were computed to give an estimate of the correlation between transcription and the H-CRAC signal (Figure 23C). In general, for both RNase H1 and H2 a good to strong correlation was found with RNAPII occupancy. Correlation with RNase H2 signal was to some

extent lower, especially for genes with a peak at their 3′-end. It is also possible that the correlation does not follow a linear model, as highly transcribed genes might generally display higher R-loop levels than expected based on the whole population. Thus, we concluded that transcription is a major driver of R-loop formation, but it does not account for all the variability observed, as expected.

Genes with a 3′-end peak of H-CRAC signal are more frequently convergent units

Beside sequence bias and transcriptional rate, topological stress is another factor that has been closely linked to R-loop formation. By visual inspection of our H-CRAC maps, we noticed that those genes having a peak at their 3′-end were often associated to a convergent transcription unit. Transcription generates positive and negative supercoiling upstream and downstream of RNAPII, respectively. Therefore, two converging RNAPIIs both introduce positive supercoils that sum up between them, and negative supercoils behind them, which can potentially lead to topological stress. In particular, negative supercoiling has been associated to R-loop formation. To test if 3′-end R-loop peaks are enriched at convergent genes, we computed the distribution of expected frequencies of convergent units using random samplings (n=10000) of equally sized groups of genes and then compared it to the number of convergent genes found with an RNase H 3′-end peak. A first attempt using the whole set of genes (795 genes) failed to find significant differences (data not shown). However, when diving the 795 in 3 group of genes according to the strength of the signal, we could observe that the first group (i.e., with the strongest signal) showed a significantly increased frequency in the number of convergent units (Figure 23D). Thus, genes with a stronger peak of H-CRAC signal at their 3′-end appear to be convergent more frequently than the ones with a weaker peak. A caveat with this analysis is that it does not include non-annotated transcription, that is frequent in the 3′-end of genes. We will consider this parameter for future analyses.

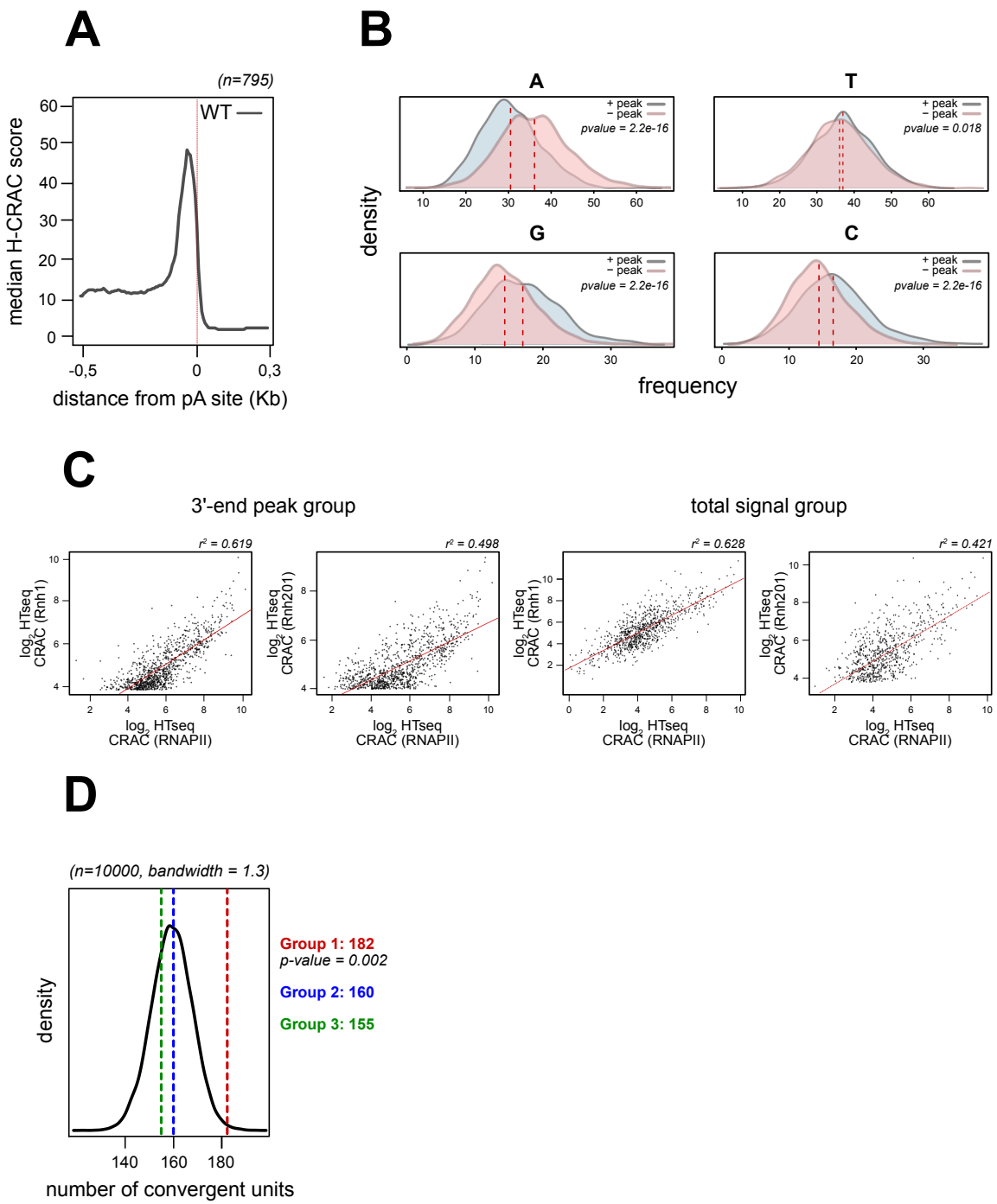


Figure 23. Transcription, nucleotide composition and topological stress likely impact RNases H binding.

A) Metagene analysis of the Rnh201 distribution at coding genes enriched for a peak of signal at their 3'-end and aligned at their pA site in wild-type (WT) cells.

B) Distributions of the nucleotide frequencies in the 100 nt region encompassing 3'-end peaks of Rnh201 (+ peak) and in the 100 nt regions centred on the pA site of an equally sized group of genes lacking a peak of Rnh201 (- peak). p-values from T tests are shown to

display the significance of the observed shifts.
C) Dispersion plots of the log_2 HT-seq value in the group of genes with a peak at their 3'-end and in the group of genes with a spread signal from H-CRAC (Rnh1) on the x axis and H-CRAC (Rnh201) on the y axis. The coefficient of determination (r^2) is shown.
D) Density of the frequencies of convergent units among randomly sampled groups of genes. The observed frequencies for the 3 group of genes with a Rnh201 peak are indicated as dashed lines. The p-value corresponds to 1 - the sum of the area left before the observation.

RNase H1 and H2 play specific roles also in physiological conditions.

RNase H2 has been proposed to have a 'housekeeping' role in R-loops clearance, while RNase H1 may serve as a back-up enzyme activated in stress conditions leading to R-loops accumulation (Lockhart et al., 2019; Zimmer and Koshland, 2016) (see Introduction, § III.IV). To gain more direct evidence for such model, we performed H-CRAC on HTP-tagged RNase H2 in cells lacking the H1 counterpart. The reciprocal experiment (Rnh1 H-CRAC in *rnh201*Δ cells) has not been performed yet but is planned for the next future.

RNase H2 plays a more prominent role than RNase H1 at coding genes

If RNase H2 plays a more prominent role, one should expect its binding to be unchanged regardless of the absence of RNase H1. Hence, we generated heatmaps of the log_2 ratio *rnh*Δ/wild-type ordered decreasingly for both the group of genes with spread signal and the one with a 3'-end peak (Figure 24A). In both cases, only a small group of genes displayed an increase in the RNase H2 signal, indicating that the absence of RNase H1 has a limited impact on the recognition of its targets by RNase H2.

Only RNase H double deletants show R-loops increased levels. This is consistent both with the notion that RNase H1 does not play major roles in physiological conditions, but fully compensates for RNase H2 absence, or with the possibility that each enzyme plays a specific role in physiological conditions, and they complement for each other

absence. To explore this last possibility, we sought for regions of the genome enriched for Rnh1 binding, and that were at the same time experiencing an increased signal of Rnh201 H-CRAC levels in *rnh1Δ* cells. Because strong differences were not observed for what concerns coding genes according to the analyses shown above, we focused on two other regions that were also shown in past studies to be prone to generate R-loops: the telomeres and the ribosomal DNA. Despite being enriched for H-CRAC signal, no noticeable differences between RNase H1 and H2 were observed at the telomeres (data not shown). Instead, at the rDNA, clear binding sites for RNase H1 were detected in the antisense to the *RDN37* gene, at the level of the Internal Transcribed Spacers 1 and 2 (*ITS1* and *ITS2*) (Figure 24C) and at the Replication Fork Barrier (RFB) (Figure 24D). Importantly, in both instances a signal for Rnh201 was also detected in wild-type cells, but it was found to be strongly increased in *rnh1Δ* cells (Figures 24C and 24D).

RNase H1 plays a more prominent role than RNase H2 at the R-loops formed in the rDNA

Thus, these data are consistent with the notion that RNase H1 and H2, although they can generally compensate for each other absence, might cooperate under physiological conditions at specific sites and eventually play specific roles. The mechanism for this specificity is not clear, but it might involve the specific recruitment of these enzymes.

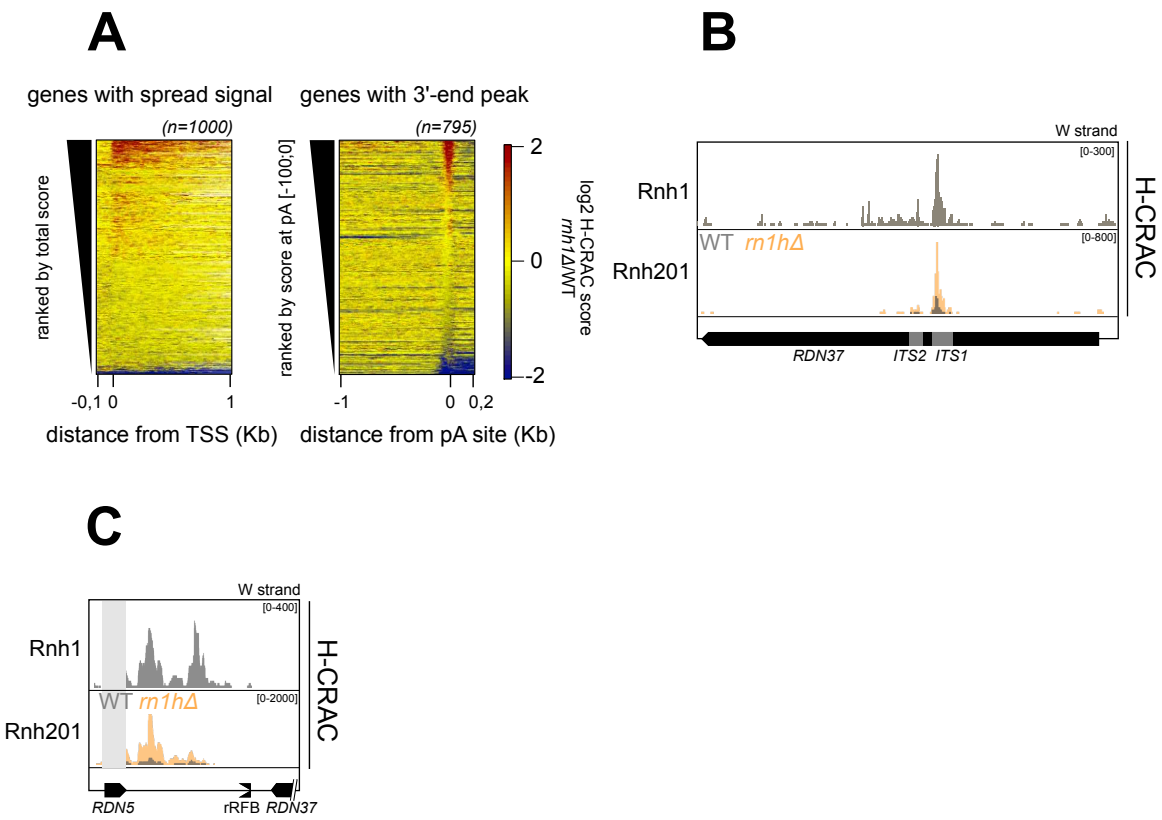


Figure 24. Rnh1 and Rnh201 play specific roles also in physiological conditions.

A) Heatmap analyses representing the $\log_2$ of the fold change (FC) of the Rnh201 signal at the two group of genes (spread signal and with a 3'-end peak) aligned to the indicated reference point. The signal of the *rnh1Δ* mutant relative to the wild-type is shown. Genes were ranked by the ratio window indicated at the left of each panel.

B) Rnh1 and Rnh201 signal antisense to the *RDN37* unit. For Rnh1 a wild-type condition is shown.

C) Rnh1 and Rnh201 signal around the rRFB. For Rnh1 a wild-type condition is shown.

Sen1 co-localises with RNases H.

The strong genetic interactions between *rnh1Δ rnh201Δ* cells and *sen1* mutants (i.e., *sen1-1*, *sen1-3*) suggests that the three proteins might have overlapping or redundant roles and possibly function at the same locations. Having already in our hands the high resolution maps of the

genomic co-localisation of RNases H and their targets, we decided to track Sen1 binding applying the CRAC methodology.

The expected binding at CUTs is detected by Sen1 CRAC

As a component of the NNS complex, Sen1 acts as a termination factor for non-coding transcription units, such as Cryptic Unstable Transcripts (CUTs). To validate our approach, we first focused on these regions. As expected, a clear signal was detected around the TES of these regions, as shown both by metagene analysis on the transcription termination sites of a group of annotated CUTs and by example cases (Figures 25A and 25B, top panel). Thus, CRAC is a valid method for detecting the binding of Sen1 to its RNA targets.

RNases H signals were also observed at CUTs, which generally follow the transcription profile, most likely as a consequence of the transcription dependency already observed at genes. We considered that in these cases the correlation between RNases H and Sen1 signals were the result of a shared dependency on transcription.

Sen1 co-localises with RNases H at the 3'-end of a common set of genes

Because strong binding of RNases H was found to occur on mRNA genes and particularly at their 3'-ends that were not explained by stronger levels of RNAPII occupancy in these regions we focused on these sites. Interestingly, and to some extent surprisingly, we found that Sen1 strongly binds to mRNA coding transcripts, as shown by heatmap analysis and example cases (Figures 25C and 25B, bottom panel). We next wondered if Sen1 and RNases H signal was limited to a common subset of genes, so we overlapped the units with the strongest scores for each dataset. As observed in Figure 25D, the three datasets strongly overlap one another, with 730 genes found in common for the tree proteins. Interestingly, Sen1 and RNase H2 displayed the greatest overlap, almost complete, with 870 genes shared between them. By visual inspection of single cases, we noticed that the pattern of the Sen1 signal along many genes was strongly reminiscent of the one observed in H-CRAC, with a marked preference for the 3'-end, peaking close to the poly(A) site (Figure 25E). This was not expected, because Sen1 is not generally involved in transcription termination of mRNA coding genes (Schaughency et al., 2014) (see Results, Chapter II). To confirm the general validity of this observation,

we computed an enrichment analysis of Sen1 binding along coding genes. In line with what we had observed by visual inspection, more than half of the signal density in mRNA genes was located at the 3'-end (Figure 25F). Moreover, heatmaps analysis centred on the pA site revealed that only a subset of genes contained a defined peak at their 3'-end (Figure 25G), mirroring what found for Rnh1 and Rnh201. Consequently, we identified the genes containing a peak of Sen1 binding at their 3'-end and assessed the overlap with the ones identifies for Rnh1 and Rnh201. Importantly, a very strong overlap was found for the three group of genes (Figure 25D). Hence, we concluded that Sen1, Rnh1 and Rnh201 co-localise at the 3'-end of a common set of genes.

The position of Sen1 at the pA site is slightly shifted compared to RNases H binding site

The three proteins bind very close to the pA site, however, their precise localisation might be slightly different, and the resolution of our methodology allows detecting even small differences. To test if this was the case, we measured the distance from the pA site of each peak for the three proteins and plotted the relative densities. Interestingly we could distinguish a shift in the position of Sen1 compared to RNases H, which were found to bind 5-10 nt before the helicase (Figure 25E).

All these observations support the idea that Sen1 and RNases H are found at the pA site of a common set of genes. Moreover, because the binding position is slightly different, it is possible to conceive that their binding is not mutually exclusive.

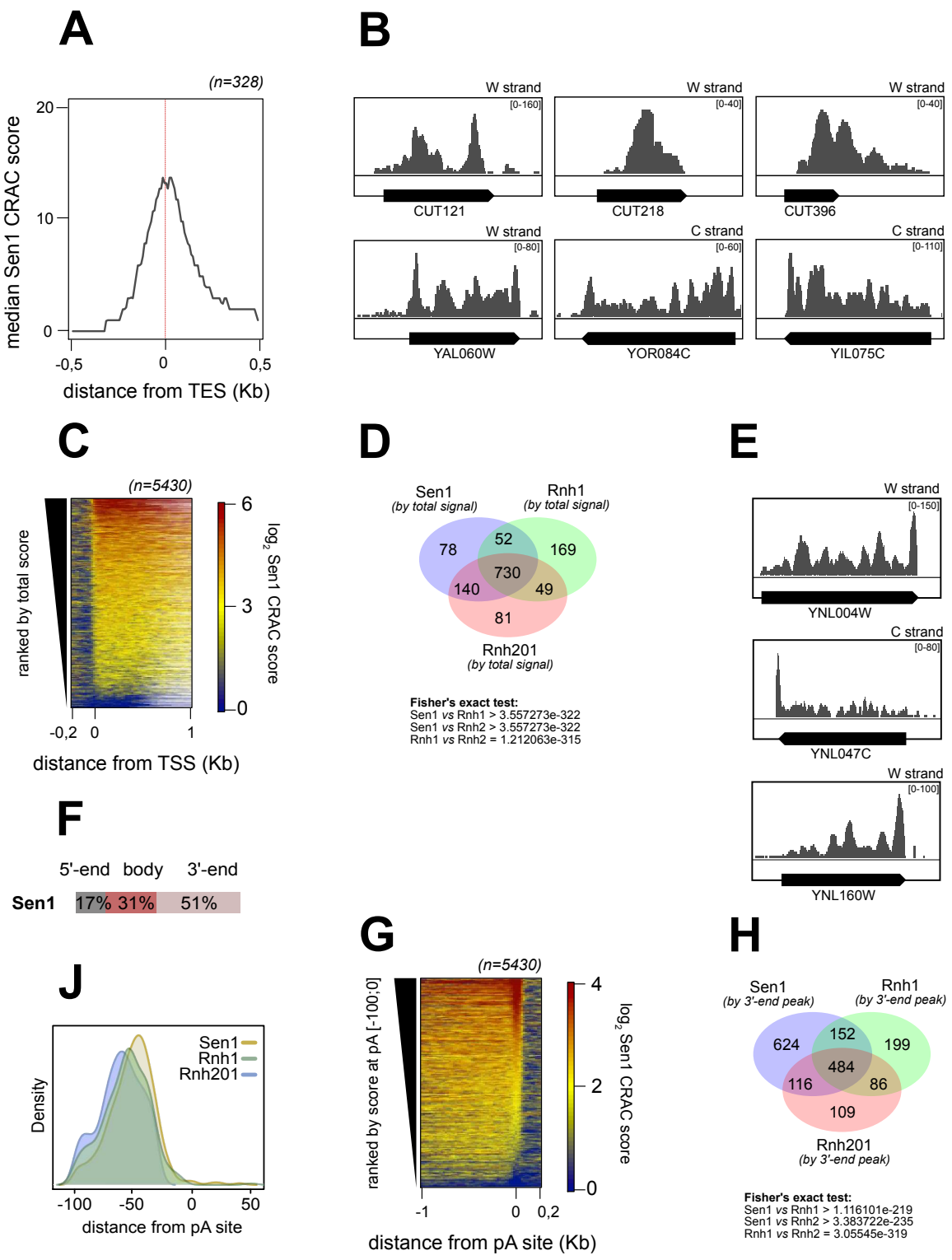


Figure 25. Sen1 co-localises with RNases H.

A) Metagene analysis of the Sen1 distribution at Cryptic Unstable Transcripts (CUTs) aligned at their TES. The signal from wild-type (WT) cells is shown.

B) Examples of Sen1 signal at CUTs and mRNA genes.

C) Heatmap analysis representing the $\log_2$ of the Sen1 signal in coding genes aligned at

their TSS in wild-type cells. Genes were ranked by the HT-seq count.
D) Overlap between the groups enriched of Sen1, Rnh1 and Rnh201 signal, selected by intensity of the signal all over the unit.
E) Examples of Sen1 signal forming a 3'-end peak in coding genes.
F) Enrichment analysis of the Sen1 signal along the 3'-end, the gene body and the 5'-end of mRNA genes. The showed values are the normalised percentages of enrichment for each region.
J) Densities of the distance of the 3'-end peak from the pA site for Sen1, Rnh1 and Rnh201.
G) As in C) but ranked by the score at the window [pA-100; pA].
H) As in D) but with genes selected for having a 3'-end peak of signal.

R-loops stabilisation stimulates Sen1 binding at protein-coding genes.

One of the possible reasons underlying the strong genetic interaction between Sen1 and RNases H is that the two enzymes might function redundantly or sequentially. For instance, it has been proposed that Sen1 resolves R-loops by virtue of its DNA:RNA helicase activity. While the increase in R-loops levels observed in *sen1-1* or Sen1-AID cells might be influenced by the termination defects linked to Sen1 misfunction, it is possible to imagine that Sen1 can unwind hybrids in mutants accumulating R-loops.

Sen1 binding at genes increases in absence of RNases H

To assess whether Sen1 binding is altered by increased R-loop levels we performed Sen1 CRAC experiments in cells lacking RNase H activity. By visual inspection of the datasets, we noticed a significant increase in Sen1 binding to its RNA targets in this condition, as shown in the examples reported in Figure 26A. To assess the significancy of this observation we compared the binding of Sen1 in *rnh1Δ rnh201Δ* and wild-type cells. As observed in Figure 26B, Sen1 levels were found to be significantly increased in *rnh1Δ rnh201Δ* cells. To determine if such increase was limited to subset of genes or rather general, we performed a heatmap analysis, which resulted in a spread and rather dispersed increase of the signal (Figure 26C). We concluded that lack of RNase H

enzyme, and likely R-loop stabilisation, stimulates Sen1 recruitment at most mRNA coding genes.

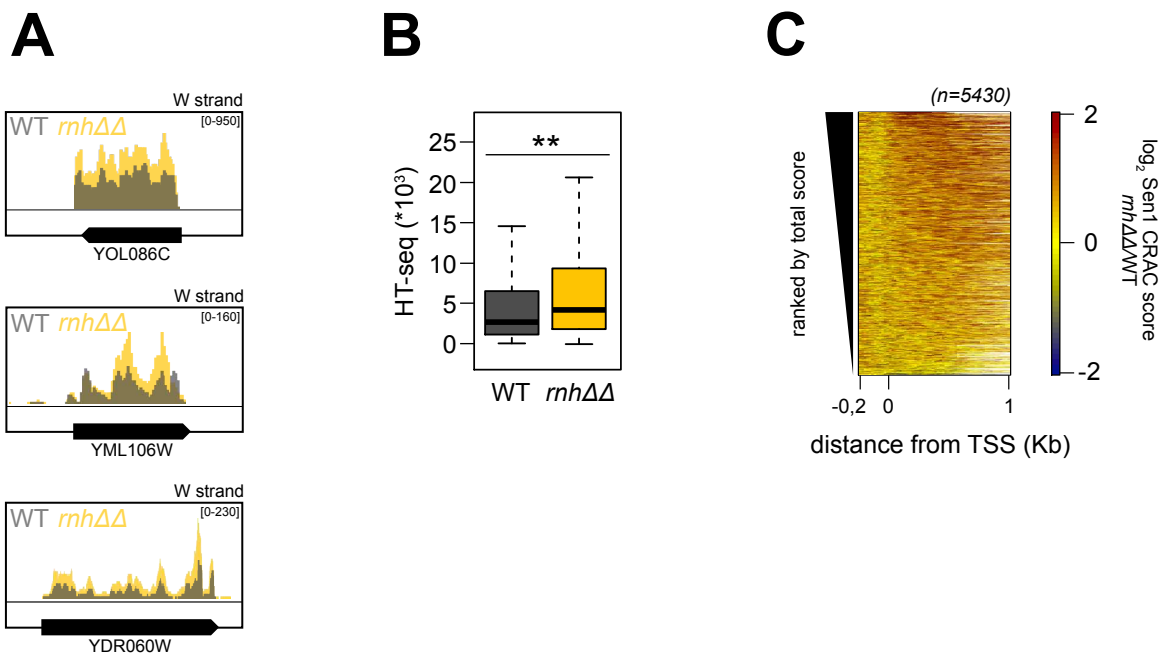


Figure 26. R-loops stabilisation stimulates Sen1 binding at coding genes.

A) Examples of increased Sen1 binding at coding genes in RNases H deleted cells (*rnhΔΔ*).

B) Comparison of the HT-seq count in coding genes between RNases H deleted (*rnhΔΔ*) and wild-type (WT) cells.

C) Heatmap analyses representing the log2 of the fold change (FC) of the Sen1 signal at coding genes aligned at their TSS. The signal of the *rnh1Δ rnh201Δ m*utant relative to the wild-type is shown. Genes were ranked by the HT-seq count.

Transcription-Replication Conflicts are not responsible for Sen1 binding at the pA site.

The co-localisation of Sen1 and RNases H at the pA site of a common set of genes was unexpected. We decided to further investigate the possible mechanistic implications of these results.

In *B. subtilis*, genes engaged in Transcription Replication Conflicts (TRCs) in a head-on conformation form strong R-loops peaking at the

transcription termination site in *rnhCΔ* cells, suggesting that RnhC promotes the resolution of high levels of R-loops that form at the TES because of TRCs (Lang et al., 2017). We hypothesised that a similar effect could be underlying the co-localisation of Sen1 and RNases H at the pA site. Therefore, we decided to perform Sen1 CRAC experiments in cells arrested in G_1 by the mean of **α**-factor addition, and so in the absence of ongoing replication. However, as observed from the heatmap performed on the group of genes containing a Sen1 peak at their 3'-end, a clear binding of the protein at the pA site could still be observed even in absence of replication (Figure 27A) similarly to what observed in the presence of replication. Thus, Sen1 binding at the pA site is not a consequence of TRCs. Analogous experiments for directly detecting R-loops by H-CRAC in G_1 have not been performed yet but have been planned.

Sen1 binds at the 3'-end even in absence of replication

We had observed that genes with a strong peak of H-CRAC signal at their 3'-ends were more frequently convergent units. Thus, we asked if this was true also for Sen1 peaks. Interestingly, we observed an even more pronounced effect, with both the group 1 (i.e., strongest binding) and the group 2 (i.e., moderate binding) showing an increased frequency of facing convergent units (Figure 27B).

Sen1 binding at the 3'-end is linked to convergent transcription

Together these observations suggest that convergent genes are more likely to have a co-localisation of Sen1 and RNase H enzymes at their 3'-ends, and we propose that this is mechanistically related to the occurrence of concurrent antisense transcription.

Sen1 interacts with Topoisomerase II and topoisomerases are required for enforcing RNAPII and Sen1 distribution at convergent genes.

An important difference between convergent and co-directional genes consists in the topological effects caused by their transcription. In co-directional genes, the positive and the negative supercoils introduced ahead and behind respectively of the transcribing RNAPII undo each

other. On the contrary, in convergent genes the positive supercoils are expected to sum-up between the two converging RNAPIIs, while negative supercoils might accumulate at the exterior borders depending on the topological arrangement of neighbouring genes. Therefore, we considered that Sen1 and RNases H might partake in the control of topological stress.

Sen1 interacts with Top2

If Sen1 binding to the pA site is related to topological stress, one could expect Sen1 to be in close proximity, or even interact, with Topoisomerases. We investigated this possibility by analysing co-Immunorecipitates (co-IP) of tagged Sen1, checking for the presence of Topoisomerases. Notably, we could not find an enrichment of Top1 compared to mock immunoprecipitation (data not shown), but a clear association between Sen1 and Top2 was detected above background (Figure 27C). Importantly, this interaction was detected in presence of nucleases treatment, although the absence of nucleic acids digestion provoked a very moderate increase in interaction (data not shown). We concluded that Sen1 interacts with Top2, directly or indirectly, but that this interaction is not DNA- or RNA-dependent.

Sen1 and RNAPII show extended read-throughs in absence of Top1 and Top2

Motivated by this observation, we moved forward and performed CRAC experiments of both RNAPII and Sen1 in cells harbouring AID-tagged versions of Top1 and Top2. Top1 and Top2 can have sometimes redundant functions, and for this reason we preferred to deplete both the two enzymes simultaneously. Visual inspection of the data produced from these experiments, revealed a surprising and completely unexpected behaviour for both Sen1 and RNAPII. As shown from example cases (Figure 27D), addition of auxin (IAA) caused a mislocalisation of Sen1, which, despite still present in the transcription unit, also localised after the pA site, with extended read-throughs (RTs) generally reaching the TSS of the following gene. This was particular evident at convergent genes, for which RTs covered completely the antisense region of the convergent unit. Importantly, both Sen1 and the polymerase followed this pattern (Figure 6D). A standard procedure to assess the level of termination defects is to compute a RT index comparing the signal retrieved after the

termination site to the one inside the transcription unit. We identified a group of 1668 showing increased RT for Sen1 (Figure 27E) and 615 for RNAPII (Figure 27F).

These results suggest the somewhat surprising conclusion that Top1, Top2 or both are required for efficient transcription termination, particularly at convergent genes that experience topological stress. Sen1 might be recruited at these sites by its interaction with Top2 and its extended signals in the absence of topoisomerases might be related to the RNAPII read-through. The interpretation of these results is, for the moment, complex, and will be briefly discussed in the following section.

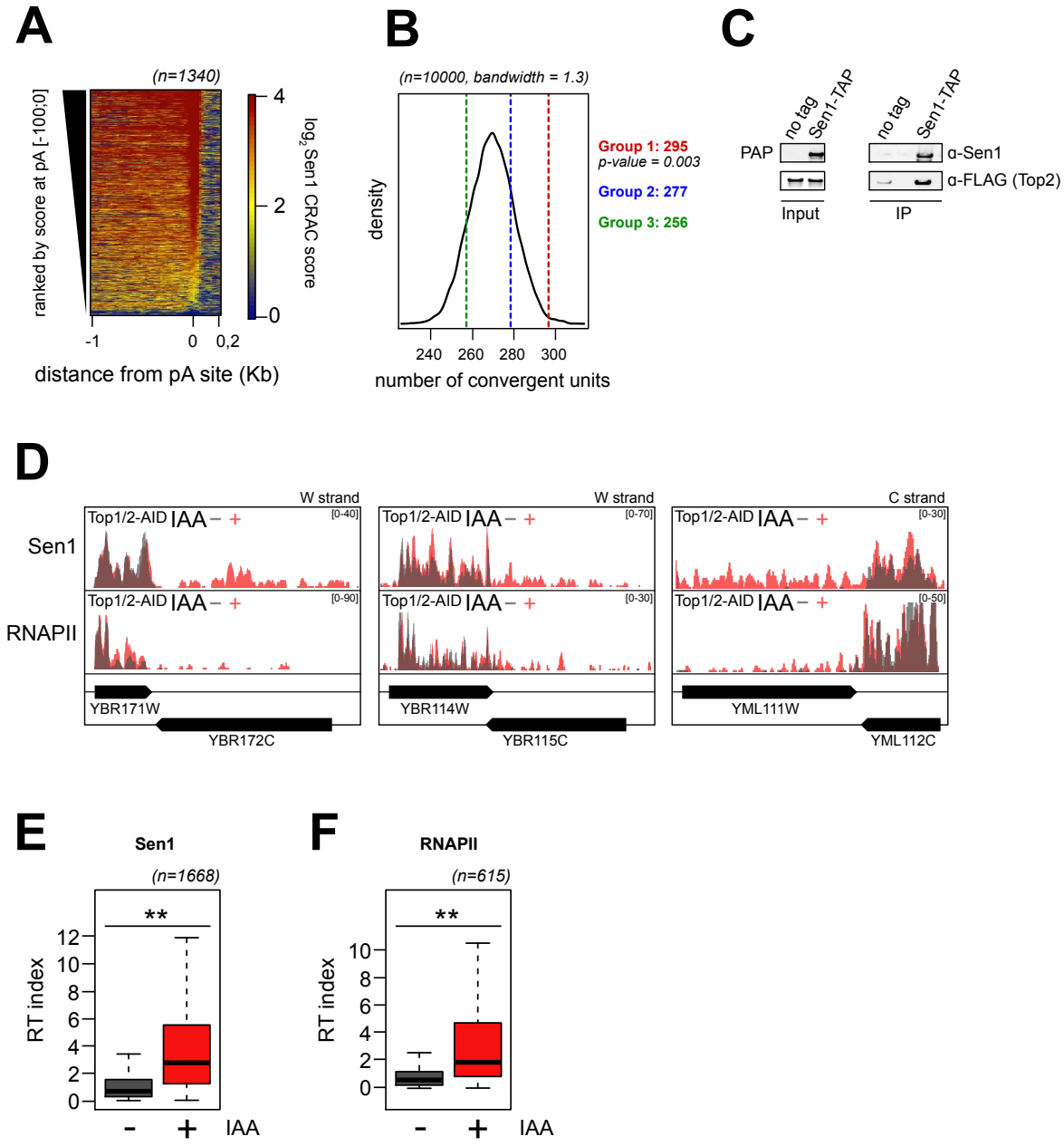


Figure 27. Sen1 interacts with Topoisomerase II

A) Heatmap analysis representing the $\log_2$ of the Sen1 signal at coding genes enriched for a 3'-end peak of Sen1 aligned at their pA site in wild-type cells arrested in G_1.

B) Density of the frequencies of convergent units among randomly sampled groups of genes. The observed frequencies for the 3 group of genes with a Sen1 peak are indicated as dashed lines. The p-value corresponds to 1 – the sum of the area left before the observation.

C) Co-IP of Sen1-TAP and analysis of the immunoprecipitates by Western Blot.

D) Examples of read-through of Sen1 and RNAPII signal at convergent genes in absence of

Top1 and Top2.

E) and **F)** RT index of Sen1 and RNAPII signal for genes selected for showing a RT in the relative dataset. **p-value<0.01

DISCUSSION AND PERSPECTIVES

Since the discovery that transcription and replication are not partitioned neither at the temporal nor spatial level, how cells coordinate these two fundamental processes and ensure that they do not interfere with each other has been a longstanding question. Moreover, the discovery that genomes are pervasively transcribed has emphasised the fact that such coordination must be even more relevant, considering the very large number of transcription events that are taking place in the genome.

Our study has focused on addressing the contribution of Sen1 and RNases H and unveiled novel functions for these proteins in the maintenance of distancing between different kind of molecular machineries. Moreover, to address some of the questions derived from this study, we have developed H-CRAC, a novel method to map R-loops genome-wide with strand specificity and high resolution.

The importance of being Sen1

Sen1 is an evolutionary conserved helicase, whose function has been thoroughly studied in the past years. As already extensively discussed, Sen1 has a fundamental function in transcription termination, and more specifically in the NNS pathway (see Introduction, § I.III.II and Results, Chapter II), but more recent studies have suggested a role for Sen1 outside the NNS complex, and more specifically in preserving genome stability.

However, all the previous reports addressing the contribution of Sen1 in genome stability have used *ts* mutants or conditional depletions of the protein, which in both cases affect all Sen1 functions, including its fundamental role in transcription termination. In this study, we have

Sen1 depletion leads to major alterations of the transcription pattern

shown that depletion of Sen1 drastically changes the cellular transcriptional landscape. We proposed that altered transcription termination at thousands of non-coding genes leads to a deep alteration of the gene expression program. Moreover, inefficient termination resulted in an increased occupancy of RNAPII in intergenic regions, which is expected to increase the chance of conflicts with replication forks. This possibility is also consistent with the notion that several factors whose mutation has been shown to lead to increased levels of R-loops or genome instability belong to the transcription termination machinery, including Pcf11, Rtt103 and Rna15 (Stirling et al., 2012). It appears very likely that none of these factors play a very direct role in the formation or resolution of R-loops/TRCs, but that their phenotypes are rather a consequence of a general tendency of altering R-loops metabolism or increasing TRCs when interfering with the well-functioning of the transcription apparatus. Consistently, *ts* mutants of Nrd1 and Nab3 also show increased Transcription-Associated Recombination (TAR) to a certain extent (Mischo et al., 2011), albeit to lower levels compared to *sen1-1* cells.

Transcription termination defects have indirect effects on R-loops and TRCs

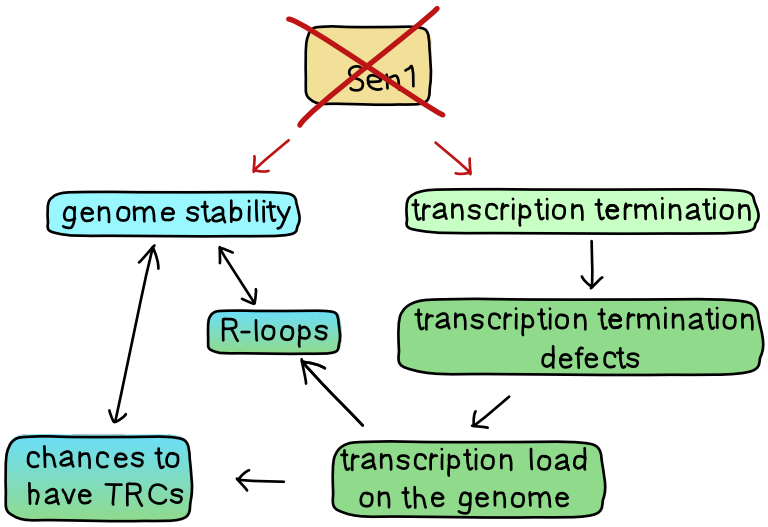


Figure 28. Sen1 roles and their interaction.

Sen1 depletion affects both the transcription termination, and the genome stability functions (red arrows). Casual relationships between the processes are indicated by black arrows and by their directionality.

The observed Sen1 phenotypes in genome instability are a combination of the multiple roles of Sen1

The marked difference between the TAR levels observed in *sen1-1* cells compared to *nrd1-102* or *nab3-11* cells (Mischo et al., 2011), as well as the fact that Sen1 only, and not Nrd1, co-localises with sites of BrdU incorporation (Alzu et al., 2012) point in favour of a *bona fide* role in genome stability. Nevertheless, we also think that the marked effect on transcription of globally defective Sen1 function cannot be neglected.

We have therefore proposed that the phenotypes that have been observed in Sen1 mutants are the consequence of *bona fide* defects in genome stability associated to defects in transcription termination, bringing to an increase in levels of R-loops or TRCs (Figure 28). Our result that coupling Sen1 absence at the replisome to Sen1-independent transcription termination defects (i.e., *sen1-3 nrd1-102*) results in levels of DNA damage comparable to the ones observed in *sen1-1* cells is in agreement with this model.

sen1-3 is a separation-of-function allele

Sen1 interacts with the replisome via its N-terminal domain

In the frame of a collaborative study, we have shown that Sen1 binds the replisome via interaction with Ctf4 and Mrc1. In a quest to map the precise site of interaction in Sen1, several truncated variants were generated, allowing the identification of a region in the N-terminal domain, whose alteration led to the creation of a new allele, named *sen1-3,* which fully loses binding to the replisome (Appanah et al., 2020). The observation that Sen1 interacts with the replisome is an important finding that support the notion that the helicase plays a role outside the NNS complex. Moreover, Sen1 is a low abundant protein (Breker et al., 2013; Ho et al., 2018), and the most widely accepted model for NNS termination proposes the helicase to be recruited to sites of action by the Nrd1 and Nab3 RNA-binding proteins (Porrua and Libri,

By interacting with the replisome, Sen1 can be recruited to sites of action

2015). Thus, the interaction with the replisome is a likely candidate for recruiting Sen1 to sites of action for its role in genome stability.

sen1-3 cells do not show transcription termination defects

The *sen1-3* allele revealed be an advantageous tool for dissecting the function of Sen1 at the replisome. On one side, *sen1-3* cells require the function of RNases H enzyme, which underlies that the recruitment of Sen1 to the replisome is mechanistically relevant for the function of the protein in maintaining genome stability. On the other hand, and differently from *sen1-1* cells or Sen1 depletion, *sen1-3* cells do not display any transcription termination defects (Appanah et al., 2020) (see Results, Chapter I), thus allowing a deeper study of the function of Sen1 at the replisome without the many indirect effects caused by an alteration in transcription termination.

Sen1 is recruited at the replisome to limit TRCs

Absence of Sen1 at the replisome causes TRCs at the 5'-end of coding genes

We have observed that the absence of Sen1 at the replisome results in an increased occupancy of RNAPII at the 5'-end of several transcription units, and exclusively in replicating cells, suggesting that the presence of the replication machinery is required for this accumulation to occur. We interpreted this observation as a consequence of missed removal of RNAPII during TRCs (Figure 29). Indeed, the density of RNAPII along transcription units is uneven, with more signal at the TSS that gradually decreases towards the TES. This could be due to an increased pausing at the beginning of the transcription cycle, as well as to futile cycles of abortive initiations, but in both cases the outcome is that the transcription machinery appears to be engaged for longer at the 5'-end compared to the 3'-end. In a pure stochastic scenario, when a replisome enters in a transcription unit it has thus more chances to encounter RNAPII at the 5'-end. Additional evidence in support of a role for Sen1 in removing RNAPII at TRCs comes from the observation that genes experiencing a stronger accumulation of RNAPII at the 5'-end in *sen1-*

3 cells are very often found to colocalise with replication forks, and that at a late replication time point (i.e., 45 min), genes that are located close to early origins, and that we determined to be replicated by Copy Number analysis, did not show such increased accumulation of RNAPII. On the same line, the fact that the genes with a higher accumulation of RNAPII change along the replication program, as observed by the lack of overlap between the affected groups at 30 and 45 min, is consistent with the notion that the position of TRCs is expected to change dynamically in the course of replication. Thus, the observed accumulation of RNAPII at the 5'-end of transcription units requires both ongoing replication and the physical co-presence at the same location of both the transcription and replication machineries.

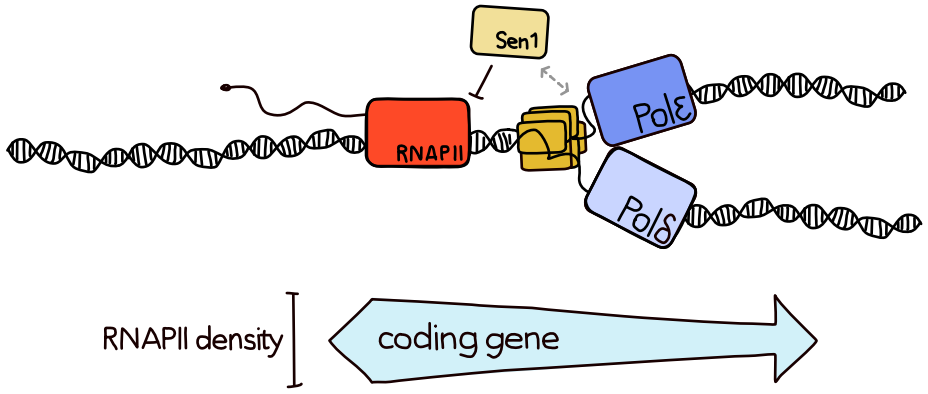


Figure 29. Sen1 releases RNAPII from TRCs at coding genes.
Schematic representation of Sen1 role at TRCs in coding genes. Sen1 is recruited to the replisome (dashed arrow) and removes RNAPII from coding units. The thickness of the coding gene arrow indicates the density of RNAPII along coding genes, and thus where TRCs are more stochastically likely to occur.

Sen1 prevents from HO and CD TRCs

A whole body of evidence demonstrated the more ill-fated outcome for head-on conflicts compared to co-directional ones (French, 1992; Hamperl et al., 2017; Lang et al., 2017). However, co-directional conflicts were also shown to occur and to induce fork stalling to a certain extent (Helmrich et al., 2013). In budding yeast, the replication fork pausing observed at RNAPII genes is independent of polarity on a genome-wide scale (Azvolinsky et al., 2009). In agreement with

previous reports, we did not detect a bias for polarity and genes with an accumulation of RNAPII were equally distributed between the head-on and co-directional groups.

Sen1 prevents from CD TRCs at the rRFB

A relevant example of co-directional TRCs is provided by the ribosomal RFB. Lack of Sen1 at the replisome provoked a marked accumulation of RNAPII at the position where the replisome is blocked by the potent polar barrier put in place by Fob1. This accumulation is located roughly 100 nt before the strongest Fob1 binding site (*RFB1*) and thus compatible with a scenario in which the replisome is stalled and RNAPII accumulates upstream and in a co-directional fashion. The existence of an unstable transcript (IGS1-R) produced by RNAPII in this region was previously reported (Houseley et al., 2007). However, it is unclear if this ncRNA corresponds to the region of transcription that we describe here, as it was detected with probes located downstream or on the rRFB, and should therefore cross the barrier. Rather, the accumulation of RNAPII that we detect occurs roughly 100 nt upstream of the rRFB. In close correspondence of this accumulation, we could detect TSSs and TESs only in a *rrp6*Δ strain, which likely define start and end of an unstable RNA located upstream of the rRFB. A function for the SETX at the rDNA was also recently reported by a study from the Mekhail laboratory. The authors have described a role for SETX in creating what they define as an R-loop shield at intergenic spacers flanking rDNA genes. The function of this shield appears to be to prevent RNAPI from producing sense intergenic noncoding RNAs that can otherwise disrupt nucleolar organization and rRNA expression (Abraham et al., 2020).

RNases H also prevent from CD TRCs at the rRFB

An important observation was that deletion of *RNH1* and *RNH201* phenocopies the effects of the *sen1-3* allele, in terms of RNAPII accumulation at the rRFB. Thus, both Sen1 and RNases H are required to prevent RNAPII accumulation at this site (Figure 30). On the contrary, lack of RNases H activity did not result in accumulation of RNAPII at the 5'-end of other transcription units. The difference between what observed at TUs and at the rRFB might be attributed to an inherent nature of this *locus* to form R-loops. Indeed, marked levels

of R-loops were shown to form at this position in the absence of Top1 and RNases H, possibly leading to a persistent transcription block (El Hage et al., 2010) and in agreement with our results. It is possible then that RNases H favour the dismantling of RNAPII, but the underlying mechanism remains unclear. A possible path to remove RNAPII would be to rely on the digestion of the RNA moiety of a hybrid to create an entry site that could be used by Rat1 to dislodge the EC with a classic torpedo termination (Figure 30). Consistently, it was recently reported that cleavage of the nascent transcript by oligonucleotide-directed RNase H can induce transcription termination (Lai et al., 2020).

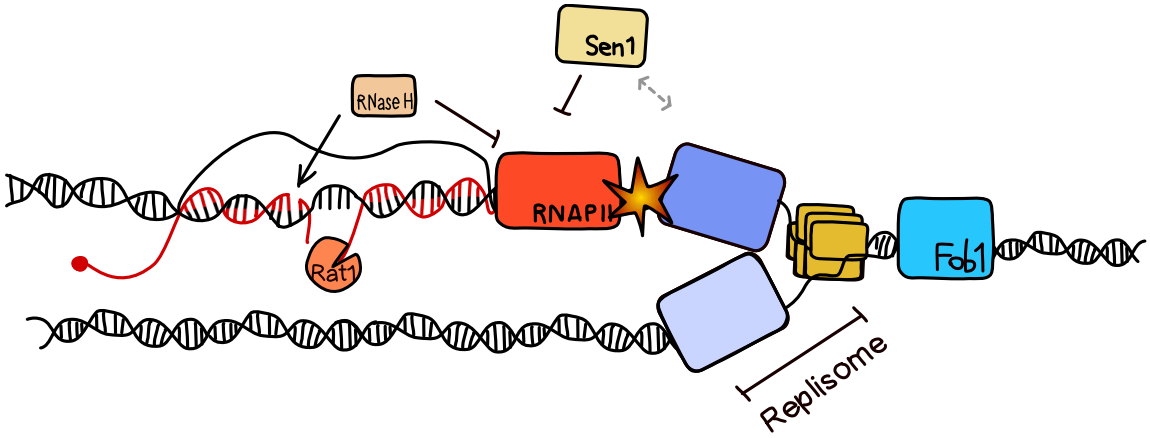


Figure 30. Sen1 and RNases H limit TRCs at the rRFB.
Schematic representation of Sen1 and RNases H roles in TRCs at the rRFB. Sen1 is recruited to the replisome (dashed arrow) and removes RNAPII. R-loop degradation by RNase H might provide an entry site for Rat1.

Do RNases H provide an entry site for Rat1?

We provided evidence of synthetic grow defects of associating the *rat1-1* and *sen1-3* alleles that is clearly stronger than the aggravation of the *rna15-2* growth defect by *sen1-3*, suggesting that it is not only a transcription defect induced by *rat1*-1 at many genes that provokes the synthetic effect with *sen1*-3. This is consistent with a role of Rat1 downstream of RNase H (therefore mimicking the effect of a double *rnh1*Δ *rnh201*Δ) but more direct evidence must be collected in the future to prove this hypothesis. A possible approach to test this hypothesis could be to track Rat1 recruitment, which could be possibly done by

CRAC, as this method was shown to be effective also on proteins engaged on the degradation of their RNA targets (Delan-Forino et al., 2017) (our H-CRAC, see Results, Chapters II and III). If cleavage of the RNA partner of an heteroduplex creates an entry site used by Rat1 for RNAPII removal, then Rat1 recruitment should be affected in RNases H depleted cells, and specifically at R-loops prone sequences.

In the frame of testing this hypothesis we undertook a different strategy, consisting in the creation of a reporter cassette that could allow the detection of transcription termination events due to such a mechanism. We have placed the R-loop prone mAIRN sequence under the control of the strong *Ptet* promoter and followed by the *CUP1 gene* under the control of a *PGAL* promoter. No terminator was introduced downstream the mAIRN sequence, so that transcription from the *Ptet* promoter would cause transcriptional interference and prevent expression of the downstream *CUP1* gene, required for growth in the presence of copper. A similar approach was successfully used in the laboratory for the selection of terminator sequences (Porrua et al., 2012). We reasoned that if the formation of an R-loop in the mAIRN region would cause RNase H cleavage and Rat1 termination, transcription interference would not occur, thus allowing expression of *CUP1*. In this perspective, in *rnh1Δ rnh201Δ* cells termination should be impaired, and mutant cells should be more sensitive to copper than wild-type cells. Despite the positive outcome in terms of generation of R-loops from the AIRN sequence (see Results, Chapter II), the reporter was very unstable in the *rnh1Δ rnh201Δ* background, with several rearrangements taking place with the general result of cutting off mAIRN expression and thus impairing the assessment of our hypothesis (data not shown). In the future, alternative approaches should rely on the sole use of wild-type cells, and a possibility would be to use several constructs with a different potential to form R-loops, so to compare them in the same wild-type background, rather than comparing the same construct in different backgrounds as we did.

This set of experiments allowed us making an interesting observation that might shed light on the mechanistic basis of RNAPII accumulation

R-loops promote *de novo* transcription initiation

at the rRFB. Indeed, by analysing the RNA molecules produced by our construct, we noticed the existence of transcripts whose 5'-end could be mapped to the position corresponding to the end of the R-loop footprint in mAIRN (Carrasco-Salas et al., 2019) (Figure 31). We have excluded that the mAIRN sequence could act as a promoter by itself because this RNA was only found when the upstream *Ptet* promoter was active (then when transcription *through* mAIRN occurred) (Figure 31A, +dox condition). Also, these products could not derive from R-loop cleavage as they were readily observed in a *rnh1Δ rnh201Δ* background (Figure 31A). Therefore, we concluded that they were the result of new transcription initiation events, which strongly suggests that R-loops can promote transcription initiation. Interestingly, a similar observation was recently made in a parallel study in mammalian cells showing that R-loops act as intrinsic RNAPII promoters and induce *de novo* antisense RNA synthesis (Tan-Wong et al., 2019).

Transposing these observations to the rRFB, a possible mechanism could be that stabilisation of R-loops by deletion of RNase H enzymes favours new transcription initiation events thus leading to increased RNAPII occupancy. The observation that *in vivo* overexpression of hsRNH1 reduced the RNAPII occupancy observed in *sen1-3* cells at the rRFB is in support of this hypothesis, even though it is also compatible with a scenario in which RNases H act instead to terminate RNAPII.

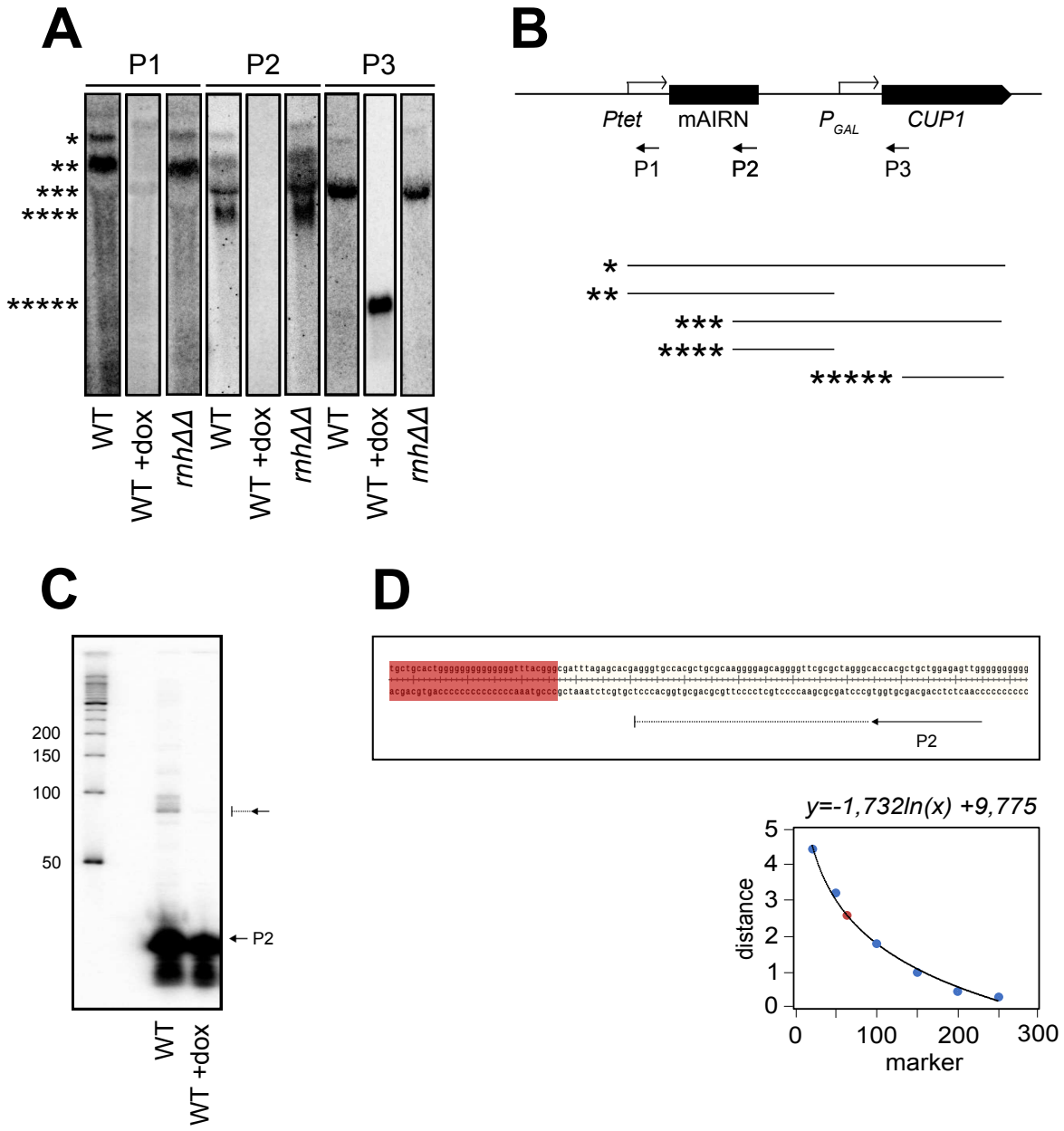


Figure 31. R-loops promote *de novo* transcription initiation.

A) Norther blot analysis of total RNAs extracted from wild-type (WT) or *rnh1Δ rnh201Δ* cells transformed with a plasmid expressing the mAIRN sequence, schematised in **(B)**, upper panel. Three different probes were used (P1, P2, P3) annealing at the indicated positions. The detected molecules are schematised in **(B)**, bottom panel.

C) Primer extension performed with the probe P2.

D) Number of nucleotides added to P2 during the primer extension showed in (C). The exact sequence where P2 is annealing is shown by an arrow, while the dashed line shows

the detected extension. The red shaded box indicated the end of the R-loop footprint according to Carrasco-Salas et al., 2019.

The beauty and the beast

We have characterised a mechanism by which Sen1, via interaction with the replisome, is brought to sites of TRCs to dislodge RNAPII and prevent from Replication Fork Pausing (RFP). However, by studying RNAPII occupancy at other genomic regions, we have also discovered that Sen1 plays a similar role also in respect to other machineries.

Sen1 prevents from RNAPII/RNAPIII conflicts

We have reported that Sen1 also interact with RNAPIII (Xie et al., 2021) and most likely using the same contact region as with the replisome, since binding to RNAPIII is also lost in *sen1-3* cells. This prompted us to monitor RNAPII occupancy nearby tRNA genes. We have observed increased RNAPII signals both upstream and antisense to tRNA genes in *sen1-3* cells. Importantly, this effect was replication-independent (i.e., observed also in G_1 arrested cells) and thus not linked to the role of Sen1 at the replisome. Thus, we reasoned that Sen1, recruited by interaction with RNAPIII, removes RNAPII that could invade tRNA genes and interfere with their transcription. This could be indeed observed in *sen1-3* cells when monitoring antisense transcription, for which it is possible to reliably detect RNAPII within RNAPIII occupancy regions.

Sen1 and RNases H prevents from RNAPI/RNAPII conflicts

A similar effect was also observed in the ribosomal unit, antisense and downstream of to the *RDN5* unit with the same modalities as for other RNAPIII TUs.

We also observed increased RNAPII occupancy antisense to the *RDN37* gene, where Sen1 also appears to remove RNAPII, possibly to avoid conflicts with RNAPI. However, the mechanistic basis for this function remains unclear because, although Sen1 also interacts with RNAPI, this interaction is not lost in the *sen1-3* mutant.

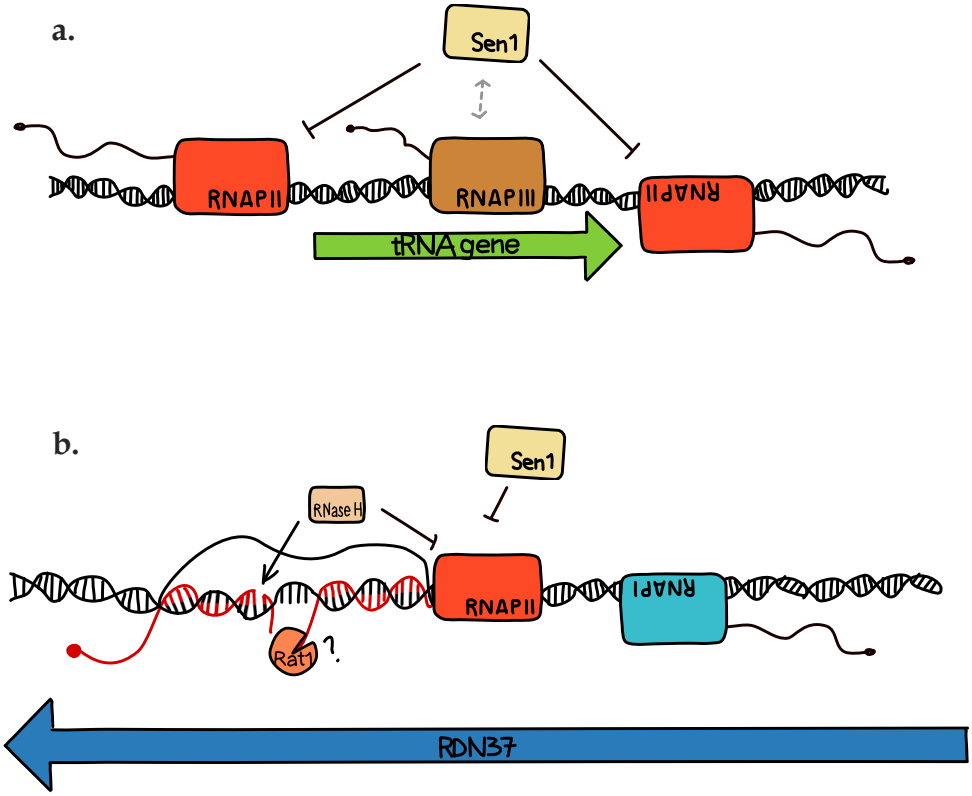


Figure 32. the role of Sen1 and RNases H in Transcription-Transcription Conflicts.
(a) Sen1 interacts with RNAPIII and limits the interferences due to RNAPII transcription by removing the EC both upstream and downstream (antisense) of tRNA genes. **(b)** Sen1 and RNases H both limit Transcription-Transcription Conflicts between RNAPII and RNAPI in the ribosomal DNA *locus*.

RPA might recruit Sen1 to the rDNA *locus*

An interesting possibility is suggested by the analysis of Sen1 co-immunoprecipitates by mass-spectrometry. Indeed, we have observed a significant decrease in the amount of recovered RPA for the *Sen1-3* protein compared to the wild-type counterpart. RPA is a ssDNA binding protein which is proposed to play a role in the recruitment of RNase H1 to R-loops (Nguyen et al., 2017). Then, Sen1 could possibly be recruited to the rDNA by interaction with RPA, considering the marked levels of R-loops formed at this *locus* (El Hage et al., 2010) (see Results, Chapter II), and remove RNAPII at conflicts with RNAPI. However, future experiments are needed to address this hypothesis,

which could be tested by studying RPA mutants, which should phenocopy the RNAPII increased occupancy at the rDNA observed in *sen1-3* cells.

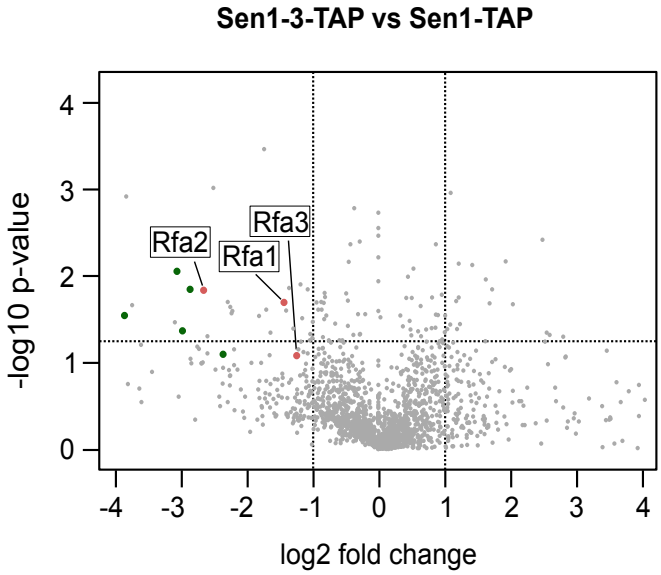


Figure 33. Sen1 might be recruited to the rDNA via interaction with RPA.
Quantitative comparison of the proteins that are associated with tagged Sen1 in a *sen1-3* mutant relative to the WT. Dashed lines are drawn at fold change = 2 and at p-value = 0.05. Green dots are replisome components.
The mass-spectrometry data presented in this figure was generated by me and it is included in another study (Xie et al., 2021).

In conclusion, we have unveiled a broader function of Sen1 in resolving conflicts between RNAPII and other machineries such as the replisome, RNAPIII and RNAPI. In full respect of its pervasive nature, RNAPII appears to be a "rude" polymerase, which does not miss the chance to transcribe anywhere it possibly can and that can only be stopped by a "caring" Sen1 that prevents it to interfere with other machineries.

Sen1 and Dicer have similar functions

A very strong similarity with Sen1 function in this sense was found for Dicer in *S. pombe*. Absence of Dicer, but not of other RNAi factors, results in the accumulation of RNAPII at coding genes, tRNA units and antisense to the rDNA *locus* (Castel et al., 2014). However, in this study

the relationships of Dicer with the replisome or with RNAPIII and RNAPI were not addressed, therefore it is impossible to conclude whether the increased RNAPII occupancy at tRNA genes or antisense to the rDNA reflects TRCs or RNAPI/III-RNAPII conflicts. Another puzzling point concern instead the position of the accumulation of RNAPII in coding genes, which was found rather towards the 3'-end in that study. The authors propose that collisions between the replication and transcription machineries first take place to the 3'-end of genes, but in my opinion, there is no clear experimental evidence for this. Transcription has a strong stochastic component, and genes are not constantly occupied by trains of polymerases, which would indeed favour collisions at the 3'-end. There is no current evidence for a mechanism that suggests coordination between the transcription and the replication machineries allowing some sort of synchronisation between the two processes ensuring that RNAPII finds itself at the end of the coding unit when the replisome approaches. Thus, the accumulation of RNAPII observed in Dicer mutants at the 3'-end as opposite to the 5'-end of TUs might be accounted for by an intrinsic difference between budding and fission yeast (i.e., the termination site in fission yeast might be a strong pausing site, where then it is more likely to find RNAPII stochastically, etc.), or to dissimilarities due to the techniques used to detect the position of RNAPII (i.e., ChIP vs CRAC).

The lethality behind a triple *sen1-3 rnh1Δ rnh201Δ*

An important point that was addressed by our study concerns the synthetic lethality between Sen1 and RNases H. We have assessed the levels of DNA damage using Rad52 *foci* formation as a readout. The single *sen1-3* allele did not show increased *foci* formation, whereas *rnh1Δ rnh201Δ* showed a marked increase, whose level was compatible with what previously observed (Amon and Koshland, 2016). The construction of an inducible triple *rnh1Δ rnh201Δ sen1-3* mutant via the use of an AID-tagged Sen1 allowed us to obtain some useful

information. For instance, the induced triple mutant displayed a significant increase in Rad52 *foci* compared to *rnh1Δ rnh201Δ* cells. These observations, together with the fact that *sen1-3* alone does not show increased Rad52 *foci* levels, overall suggest that the lethality is likely to be due to an essential need of Sen1 at the replisome in *a rnh1Δ rnh201Δ* background, rather than the other way around. A possible interpretation is that stabilisation of R-loops makes somehow TRCs more likely and thus Sen1 at the replisome is required with more frequency than what normally required. Another and not mutually exclusive possibility is that it is the lack of co-function of Sen1 and RNases H in conflicts at the rRFB to lead to extensive DNA damage. A simple way to test this hypothesis would be to assess whether the synthetic lethality of the *rnh1Δ rnh201Δ sen1-3* can be suppressed in a *fob1Δ* background. Because Fob1 is essential to establish the rRFB, a prediction would be that its lack would cancel TRCs at this position and consequently the generation of the associated DNA damage, and possibly lethality.

Fantastic methods and where to find them

The study of R-loops has received great attention in the recent years by the scientific community. However, currently used methodologies to assess R-loops levels and distribution suffer from sometimes marked inconsistencies. The crosslinking of RNases H to its targets in H-CRAC has revealed to be a very effective method. We have successfully generated maps of Rnh1, Rnh201 and hsRNH1 targets and showed that they largely overlap. We also have implemented several controls to ascertain the reliability of our approach, which all resulted in positive outcomes confirming that H-CRAC is a valid method to map R-loops. Opposite to ChIP based methodologies, H-CRAC offers an unprecedented resolution which allowed us to make some interesting observations. The only current other method granting a resolution

comparable to the one of H-CRAC is SMRF-seq, which also allows the profiling of single R-loops. To our knowledge this method has not been applied to yeast. On the other side of the medal, SMRF-seq is not suitable for genome-wide analysis at the moment, while H-CRAC was specifically designed as a genome-wide method.

Rnh1, Rnh201 and hsRNH1 all bind at the pA site

Preliminary analyses of our H-CRAC maps reaved prominent binding of all tested RNases H at the 3' end of genes, close to the pA site. Formation of R-loops in a similar position was already observed in human cells (Ginno et al., 2013; Promonet et al., 2020) and claimed to be observed in yeast albeit the resolution and quality of the generated data is too poor for reliable conclusions (Achar et al., 2020). The presence of a signal near the TES is one of the main discrepancies between S9.6-DRIP-seq and R-ChIP, with the latter generally depleted of a signal at termination regions that it is instead found by the former. Thus, H-CRAC reconciles for the first time what observed with the two approached (i.e., S9.6 and RNases H binding). In the future it will be very informative to perform H-CRAC in human cells, the only model for which enrichment of the signal at the 3'-end is clearly observed, and test if also in this case a binding of RNases H is observed at this position.

The mechanistic basis of RNases H binding at this potion are rather unclear. While in humans, R-loops were proposed to play a role in transcription termination (Skourti-Stathaki et al., 2011; Zhao et al., 2016), such a link is rather missing in budding yeast. However, some interesting hypothesis can be proposed in light of the results that were found (see below).

RNase H2 takes it all, almost

The specific function of RNase H enzymes has been a difficult point to address, because of a great deal of overlap and/or redundancy between the two enzymes. The only undisputed specificity concerns the removal

of single rNMPs that are erroneously incorporated during DNA synthesis and that can only be fulfilled by RNase H2.

RNase H2 is the 'housekeeping' enzyme for R-loops degradation

For what concerns instead R-loops degradation, the idea that one or both the enzymes function to digest hybrids is supported by strong evidence, such as the increased accumulation of R-loops in *rnh1Δ rnh201Δ* cells, or their decrease when overexpressing hsRNH1. Still, the specificity of these two enzymes remains in the shade. Two different studies (Lockhart et al., 2019; Zimmer and Koshland, 2016) with different approaches agreed on a model that proposes RNase H2 to play a 'housekeeping' role in the resolution of R-loops in physiological conditions, while RNase H1 would function as a back-up enzyme and enter in action only at particularly persistent R-loops. By analysing our H-CRAC maps we have found that RNase H2 binding is mostly unchanged in a *rnh1Δ*, which can be expected in a scenario in which RNase H1 has a generally auxiliary function, and thus in line with previous studies.

Rnh1 and Rnh201 act at distinguished genomic regions

Nevertheless, our work is the first to address the genome-wide binding of both enzymes and its precision should allow addressing the specificities of the two enzymes. For instance, we observed a clear binding of Rnh1 at R-loops formed at the rDNA *locus*, both at the rRFB and antisense to the *RDN37* gene, while binding of RNase H2 at these positions was rather low in a wild-type context, but strongly increased in *rnh1Δ* cells. The preliminary analysis of our data generally supports the notion that RNase H2 has a predominant role in R-loops digestion, because of a stronger and generally more extended binding. However, care should be taken in interpreting these findings as stronger binding might result from many factors, including differences in crosslinking efficiency and protein extraction. An important, missing, but planned, experiment is to perform Rnh1 H-CRAC in *rnh201Δ* cells, which we expect to result in a strong Rnh1 signal increase compared to the wild-type control, and could provide the missing piece of evidence to formally support the proposed model.

Determinants of RNases H binding

Nucleotide composition and transcription rate influence H-CRAC signal

Not all genes form R-loops, and among those forming some, much difference is found in terms of their specific levels. What makes a gene more prone to form an R-loop compared to another? Several studies have already identified features that underlie the observed differences between R-loops prone and R-loop depleted genes. We have performed preliminary analyses of our H-CRAC maps aiming to test if RNases H targets followed the expected behaviour based on previous findings and observed some agreement with the literature. For instance, we found a strong correlation with transcription rate, which explains a large fraction of the observed variability in H-CRAC signal. We also observed a bias in nucleotide composition for the genes forming a peak of RNase H at their 3'-end, in line with what observed in human cells (Ginno et al., 2013). However, some interesting observation that requires additional investigation were also made. In particular, we have found a weaker correlation between transcription rate and RNase H2 binding, especially for those genes forming a peak at the pA site. In the future it will be opportune to investigate what differentiates those genes with an equal transcription rate but different Rnh201 levels.

An interesting feature that we found to influence RNases H binding, and particularly when the 3'-end peak is observed, was the presence of a convergent transcription unit, which might relate to the different topological constraints that convergent genes are expected to experience compared to co-directional or divergent genes.

The mysterious affair at the pA site

An important and unexpected finding was the co-localisation between Sen1 and RNases H enzymes upstream of the pA site. The basis of this

Binding of Sen1 and RNases H to the pA might be linked to topological stress

co-localisation remains still not understood, but the observation that convergent genes are more likely to have a peak of Sen1/RNases H binding opens the window for interesting speculations and connections with topological stress. The relevance of the topological changes due to transcription was already addressed in the past by observing the effect on transcription caused by the inhibition of topoisomerases at the rDNA *locus* (French et al., 2011), leading to a model that proposed the positive torsion in front of the polymerase to be largely resolved by Top2, whereas the negative torsion behind the polymerase to be dispersed by Top1. A study from the Aguilera lab specifically focused on the need of topoisomerases activity for convergent transcription, although in a completely artificial system (García-Rubio and Aguilera, 2012).

One example provides interesting clues that convergent transcription is required for Sen1 localization at convergent genes. *HSP150* (YJL159C) is a ~2Kb long gene induced by the heat shock response but nevertheless transcribed also in physiological conditions and regardless of the cell cycle phase. It is located on the W strand of the left arm of ChrX and it is immediately followed by a convergent gene, *CIS3* (YJL158C) encoded on the C strand. *CIS3* transcription is cell cycle regulated, being the gene repressed in G_1 arrested cells. Interestingly, Sen1 forms a prominent peak at the pA site of *HSP150*, but this peak is strongly decreased in G_1 arrested cells despite the fact that the transcription levels of the gene are unchanged. Thus, activation of the convergent gene appears to be required for Sen1 recruitment at *HSP150* pA site. Notably, also *CIS3* is decorated with a Sen1 peak at his pA in G_1 when it is actively transcribed.

Sen1 interacts with Top2

In the frame of better understanding the connection between Sen1, the pA peak and the link to topological stress, we have discovered a physical interaction between Sen1 and Top2 and obtained the unexpected result that lack of both Top1 and Top2 results in a mislocalisation of Sen1 and extended read-through of RNAPII at convergent genes. In the future it will be important to ensure that the increased occupancy of RNAPII antisense to convergent gene is a result

of termination defects and not of new transcriptional events, which can easily be assessed by analysing the produced RNAs by Norther blot.

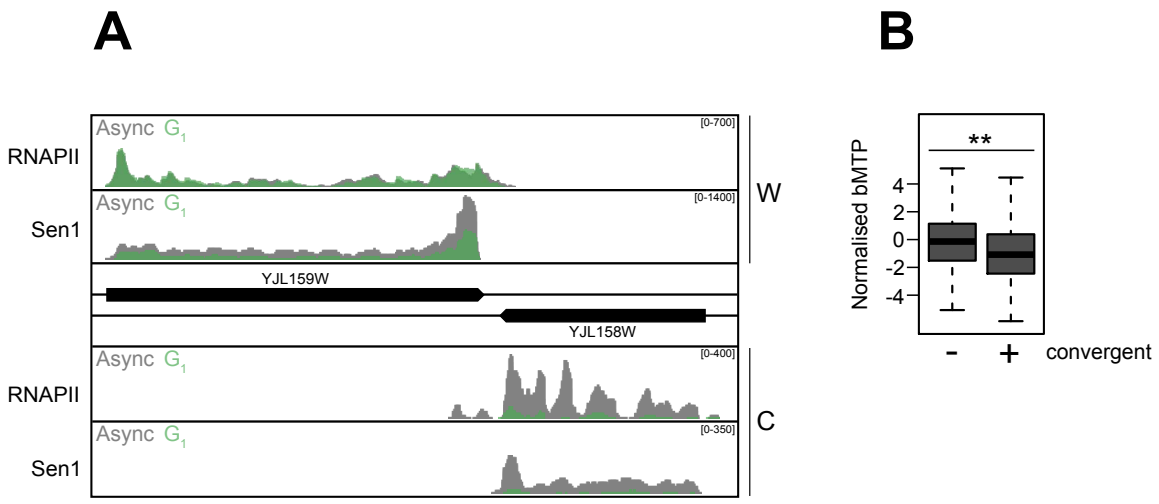


Figure 34. Topology might be linked to the formation of a Sen1 peak at the 3'-end of coding genes.

A) The example case of *HSP150*.

B) Comparison of the topology between convergent (+) and not convergent (-) genes.

Does topology have an influence on transcription termination?

If confirmed, these findings might suggest a connection with transcription termination. An interesting difference between convergent and co-direction genes concerns their termination region. In co-directional genes the termination region of the upstream gene is followed by the promoter of the downstream gene. A direct consequence of this genomic arrangement is that transcription factors bound at the promoter of the downstream gene act as "roadblocks" to terminate RNAPII failing to terminate at the upstream terminator, in order to protect the downstream promoter from transcriptional interference (Candelli et al., 2018; Colin et al., 2014). In convergent genes such protection is obviously not ensured and the polymerases that fail to terminate at one of the two gene might enter in the antisense region of the convergent unit.

One of the direct consequences of increased torsional stress is to slow down the EC. Thus, an exciting possibility could be that a certain level of positive torsional stress must be maintained at the termination region of convergent genes to pause RNAPII and facilitate their termination. We have verified the existence of this positive supercoiling at the termination regions of convergent units by analysis a recently published dataset of topology distribution genome-wide (Achar et al., 2020) realised using 4,5,8-trimethylpsoralen (TMP) intercalation to monitor negative superhelical tension. By comparing a group of genes that have a convergent unit to an equivalent group lacking it, we could observe a significant enrichment of positive supercoils at the pA site.

However, excess positive supercoiling, as expected to occur between convergent genes in the absence of topoisomerases, would rather inhibit transcription termination. This is counterintuitive to some extent, but maybe excess supercoiling alters the binding of proteins to the DNA region downstream of the TES and allows extended readthroughs that occur constitutively, but are normally limited in intergenic regions.

Sen1 could be recruited or maintained at these sites by its interaction with Top2, but its role remains unclear as transcription termination defects have not been detected at these sites even upon its full loss of function or depletion (Schaughency et al., 2014). Another function must be envisaged, which might be related to the necessity for unwinding R-loops or removing RNAPII that form excess R-loops. The role of RNases H and the impact of R-loops in these regions is also unclear. R-loops might form because of the negative supercoiling, and hinder additional transcriptional events unless removed by RNases H. It is also possible that Sen1 is recruited to this region (maybe via Top2) to remove polymerases that form excess R-loops, thus providing a redundant role with RNases H. These possibilities, although very exciting, remain for the moment mostly based on speculations, and more direct observations are needed to test them. In the future, possible approaches to tackle the question could rely on the use of plasmid-borne constructs specifically designed to study the binding of Sen1/RNases H at

convergent genes, eventually modifying the orientation, distance and occurrence of convergent transcription. Whether formation of these peaks require the commitment to transcription termination should also be assessed, using mutants in the CPF pathway, and/or removing transcription termination signals. Cohesins have been shown to be "pushed" by transcription in intergenic regions (Lengronne et al., 2004). Binding of these proteins might be affected in the presence of excess supercoiling, when topoisomerases are depleted, and they might be the barriers that prevent transcription leaking from terminators to invade convergent genes.

On the short term time scale, we hope to deepen our understanding by a more thorough analysis of the data we already produced, which, for lack of time, have undergone only preliminary scrutiny.

MATERIALS AND METHODS

Yeast strains and plasmids and standard genetics

Saccharomyces cerevisiae is the experimental model used in this study. All used strains are isogenic to BMA64 and are listed in Key Resource Table, as well as used plasmids. Construction of new strains and plasmids was conducted with standard procedures (Longtine et al., 1998).

Cell growth for CRAC

For each condition, 2L of cells expressing an HTP-tagged version of the protein of interest (i.e., Rpb1, Sen1) from the endogenous locus were grown in logarithmic phase until OD600=0.6 at 30oC in conditional medium.

All cultures were constantly checked for contaminations by visualisation at the microscope during cell growth

For conditions arrested in G1, at OD600=0.3 cell cycle arrest was triggered by 3 consecutive additions at interval of 40m of 4, 8 and 4 mg of α-factor, respectively. 40m after the last addition of α-factor, and before cross-linking, proper arrest was systematically checked by both visualisation of cell morphology at the microscope and by FACS analysis.

For (+) IAA conditions, Indole-3'-Acetic Acid (Sigma) was supplemented at a final concentration of 500µM 1h before cross-linking.

UV-Crosslinking and cDNA analysis (CRAC)

The CRAC protocol used in this study is already described in Results, Chapter II.

For Sen1 CRAC some modifications from were made to improve the recovery of the tagged protein of interest. In particular, DNase I treatment was replaced by a step of chromatic shredding submitting thawed lysates to a cycle of sonication in an ice-cold bath (15 min, High, 45 sec ON/OFF, Diagenode). The GelFree fractionation was skipped because of lack of contaminants after the second purification step and to avoid further loss of material.

Co-Immunoprecipitation (Co-IP)

For immunoprecipitation cells harbouring a tagged Sen1 expressed under its own promoter were grown in exponential phase on YPD medium. Cultures (typically 250 mL) were grown to OD600=1 and then cells were collected by centrifugation and resuspended in 1.5 mL of lysis buffer (10 mM sodium phosphate pH 7.5, 200 mM sodium acetate, 0.25 % NP-40, 2 mM EDTA, 1 mM EGTA, 5% glycerol) containing fresh protease inhibitors (2mM AEBSF, 1mM benzamidine, EDTA-free anti-proteases tablet). Suspensions were frozen in liquid nitrogen and lysed using a Retsch MM301 Ball Mill (5 cycles of 3 minutes at 15 Hz). Lysates were supplemented with 150 U/mL of lysate of Pierce Universal Nuclease (Thermo Fisher) and incubated for 40 min at 4°C for nucleic acids digestion and then clarified by centrifugation at 13 krpm for 30 min at 4°C. The extracts were then incubated with 2.5 mg of IgG-coupled M-280 tosylactivated dynabeads (Thermo Fisher) for 2 h at 4°C with rotation. After incubation, beads were washed four times with lysis buffer and once with TEV cleavage buffer (10 mM Tris-HCl, 150 mM NaCl, 0,1 % NP40, 0.5 mM EDTA, 5 % glycerol, 1 mM DTT). Elution was performed by cleaving with 40 U of TEV protease for 1.5 h at 25°C in a final volume of 120 µL.

Inputs and eluates were incubated at 95°C for 5 m with commercial Leammli loading buffer (Biorad) and then analysed by standard immunoblotting procedures. Detection was performed using SuperSignal West Pico Chemiluminescent Substrate (ThermoFischer Scientific) and a ChemiDoc Imaging system (Biorad).

RNA analysis

RNAs were prepared by the hot acid phenol extraction. Briefly, yeast cells were spinned and resuspended in 400µl of AE buffer (50 mM Sodium acetate (pH 5.5), 10 mM EDTA, 1% SDS). An equal volume of water-saturated phenol was added, and the samples were incubated for 30min at 65°C with shaking. The aqueous phase was recovered. The phenol extraction was repeated once with water saturated phenol and once with chloroform. The RNAs were ethanol precipitated. For Northern blot analysis, 10 µg of RNA were separated by agarose gel electrophoresis and transferred to a Hybond Nylon N+, membrane (GE Helthcare) by capillarity. Hybrydization was performed in UltraHyb buffer (Ambion).

Dataset processing and data analysis

The H-CRAC and some of the RNAPII CRAC maps were obtained in the frame of a previous study (see Results, Chapter II).

CRAC

Raw CRAC datasets were analysed as previously described (Candelli et al., 2018; Challal et al., 2018). The pyCRAC script pyFastqDuplicateRemover was used to collapse PCR duplicates using a 6 nucleotides random tag included in the 3′ adaptor (see Key Resources Table). The resulting sequences were reverse complemented with Fastx reverse complement (part of the fastx toolkit, http://hannonlab.cshl.edu/ fastx_toolkit/) and mapped to the R64 genome (Cherry et al., 2012) with bowtie2 (-N 1) (Langmead and Salzberg, 2012).

Quantification and statistical analysis

All analysis were performed with inhouse scripts in the RStudio environment.

Enrichment analysis along coding units was performed by computing the scores at the 5′-end [TSS; TSS+100], in the body [TSS+100; pA-100] and at the 3′-end [pA-100; pA]. All the scores were then normalised to their respective sizes and a percentage of enrichment for each position was calculated and plotted with scaled bars.

Nucleotide frequency was computed on a 100nt window centred on the highest value of each peak. A group of equally sized genes without a 3′-end peak of Rnh2 signal was used as comparison. T tests were used for the statistical analysis.

Assessment of convergency for genes with a 3′-end peak was evaluated by first performing Monte Carlo simulation (N=10000) to obtain a distribution of frequencies of convergent units for random groups of genes. A unit was considered convergent when in the window [pA-100; pA+500] a second gene was found to overlap in the opposite strand. The p-values were calculated as 1 - the density left behind the observed point.

Where appropriate, data was presented as the average with error bars representing standard deviation. t tests were used to compare population means. Statistically significant differences were indicated as such by indicating the value range of the p values.

KEY RESOURCE TABLE

Antibodies		
Reagent or Resource	**Source**	**Identifier**
IgG from rabbit serum	Sigma-Aldrich	Cat# I5006; RRID: AB_1163659
Mouse anti Flag	Sigma-Aldrich	Cat# F1804; RRID: AB_262044
Rabbit Peroxidase Anti-Peroxidase	Sigma-Aldrich	Cat# P1291; RRID: AB_1079562
Goat anti-rabbit IgG-HRP	Santa Cruz	Cat# sc-2004; RRID: AB_631746
Goat anti-mouse IgG-HRP	Santa Cruz	Cat# sc-2005; RRID: AB_631736

Chemicals, Peptides, and Recombinant Proteins		
Reagent or Resource	**Source**	**Identifier**
cOmplete EDTA-free protease inhibitor cocktail tablets	Sigma-Aldrich (Roche)	Cat# 11873580001
Pefabloc SC-Protease-Inhibitor	Carl Roth	Cat# A154.3
DNase I recombinant, RNase-free	Sigma-Aldrich (Roche)	Cat# 04716728001

Dynabeads M-280 Tosylactivated	Thermo Fisher Scientific	Cat# 14204
Recombinant GST-TEV protease	(Challal et al., 2018)	N/A
RNace-It Ribonuclease Cocktail	Agilent	Cat# 400720
Guanidine hydrochloride	Sigma-Aldrich	Cat# G4505
Ni-NTA Agarose	QIAGEN	Cat# 30230
Imidazole	Sigma-Aldrich	Cat# I0125
RNaseOUT Recombinant Ribonuclease Inhibitor	Thermo Fisher Scientific	Cat# 10777019
T4 RNA Ligase 2, truncated KQ	NEB	Cat# M0373L
T4 Polynucleotide Kinase	NEB	Cat# M0201L
T4 RNA Ligase 1 (ssRNA Ligase)	NEB	Cat# M0204L
Proteinase K, recombinant, PCR grade	Sigma-Aldrich (Roche)	Cat# 03115887001
SuperScript IV Reverse Transcriptase	Thermo Fisher Scientific	Cat# 18090050
RNase H	NEB	Cat# M0297S
Exonuclease I	NEB	Cat# M0293S
LA Taq	Takara	Cat# RR002M
α-factor	BIOTEM	N/A
Pierce Universal Nuclease	Thermo Fisher Scientific	Cat# 88700
3-Indoleacetic acid (IAA)	Sigma-Aldrich	Cat# I2886

Strains		
Reagent or Resource	**Source**	**Identifier**
DLY671 wild-type	F. Lacroute	as BMA64; Mat a
DLY2571 Rpb1-HTP	(Candelli et al., 2018)	as BMA64; RPB1::HTP::TRP1kl; Mat a
DLY3217 Sen1-HTP	This study	As BMA64; SEN1::HTP::TRP1kl; Mat a
DLY3412 Sen1-HTP rnh1Δ rnh201Δ	This study	as BMA64; SEN1::HTP::TRP1kl; rnh1::hphNT; rnh201::HisMX; Mat a
DLY3508 Top2-3xFLAG	This study	as BMA64; TOP2::3xFLAG;; NatMx; Mat a
DLY3515 Sen1-TAP Top2-3xFLAG	This study	as BMA64; SEN1::TAP::KanMX; TOP2::3xFLAG;;NatMx; Mat a
DLY3516 Rpb1-HTP Top1-AID Top2-AID	This study	as BMA64, RPB1-HTP::TRP1kl, ura3-1::pADH1-OsTIR1-URA3, TOP2-AID-MYC-HPH TOP1-AID-KanMX6 Mat a
DLY3538 Sen1-HTP Top1-AID Top2-AID	This study	as BMA64, SEN1-HTP::TRP1kl, ura3-1::pADH1-OsTIR1-URA3, TOP2-AID-MYC-HPH TOP1-AID-KanMX6 Mat a

Critical commercial assay		
Reagent or Resource	**Source**	**Identifier**
LightCycler 480 SYBR Green I Master	Roche	Cat# 04887352001
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel	Cat# 740609
Pierce Spin Columns - Snap Cap	Thermo Fisher Scientific	Cat# 69725
Vivacon 500	Sartorius	Cat# VN01H22
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific (Invitrogen)	Cat# Q32851
SuperSignal West Pico Chemiluminescent Substrate	Thermo Fisher Scientific	Cat# 34080

Software and Algorithms		
Reagent or Resource	**Source**	**Identifier**
RStudio	RStudio	RRID:SCR_000432
ImageJ	NIH	https://imagej.nih.gov/ij/; RRID:SCR_003070
Affinity Designer	Serif	https://affinity.serif.com/en-us/designer/; RRID:SCR_016952

Oligonucleotides		
Reagent or Resource	**Source**	**Identifier**
DL378- ACACTTGTGGTGAACGATAG	This study	N/A
DL751- TTTCCCAGAGCAGCATGACT	This study	N/A
DL1522- GAGCCCTTTCTGTAAATTGC	This study	N/A
DL4520- CAACTCTCCAGCAGCGTGGT	This study	N/A

REFERENCES

Abe, T., Sugimura, K., Hosono, Y., Takami, Y., Akita, M., Yoshimura, A., Tada, S., Nakayama, T., Murofushi, H., Okumura, K., et al. (2011). The histone chaperone facilitates chromatin transcription (FACT) protein maintains normal replication fork rates. J. Biol. Chem. *286*, 30504–30512.

Abraham, K.J., Khosraviani, N., Chan, J.N.Y., Gorthi, A., Samman, A., Zhao, D.Y., Wang, M., Bokros, M., Vidya, E., Ostrowski, L.A., et al. (2020). Nucleolar RNA polymerase II drives ribosome biogenesis. Nature *585*, 298–302.

Achar, Y.J., Adhil, M., Choudhary, R., Gilbert, N., and Foiani, M. (2020). Negative supercoil at gene boundaries modulates gene topology. Nature *577*, 701–705.

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

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

Ahn, S.H., Kim, M., and Buratowski, S. (2004). Phosphorylation of Serine 2 within the RNA Polymerase II C-Terminal Domain Couples Transcription and 3′ End Processing. Mol. Cell *13*, 67–76.

Alberts, B.M., Barry, J., Bedinger, P., Formosa, T., Jongeneel, C.V., and Kreuzer, K.N. (1983). Studies on DNA replication in the bacteriophage T4 in vitro system. Cold Spring Harb. Symp. Quant. Biol. *47 Pt 2*, 655–668.

Allen, J.B., Zhou, Z., Siede, W., Friedberg, E.C., and Elledge, S.J. (1994). The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. Genes Dev. *8*, 2401–2415.

Alzu, A., Bermejo, R., Begnis, M., Lucca, C., Piccini, D., Carotenuto, W., Saponaro, M., Brambati, A., Cocito, A., Foiani, M., et al. (2012). Senataxin Associates with Replication Forks to Protect Fork Integrity across RNA-Polymerase-II-Transcribed Genes. Cell *151*, 835–846.

Amon, J.D., and Koshland, D. (2016). RNase H enables efficient repair of R-loop induced DNA damage. ELife *5*, e20533.

Amrani, N., Minet, M., Le Gouar, M., Lacroute, F., and Wyers, F. (1997). Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. Mol. Cell. Biol. *17*, 3694–3701.

Anupama, K., Leela, J.K., and Gowrishankar, J. (2019). Two pathways for RNase E action in Escherichia coli in vivo and bypass of its essentiality in mutants defective for Rho-dependent transcription termination. Mol. Microbiol. *112*, 1371.

Aparicio, O.M., Weinstein, D.M., and Bell, S.P. (1997). Components and Dynamics of DNA Replication Complexes in S. cerevisiae: Redistribution of MCM Proteins and Cdc45p during S Phase. Cell *91*, 59–69.

Appanah, R., Lones, E.C., Aiello, U., Libri, D., and De Piccoli, G. (2020). Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability. Cell Rep. *30*, 2094-2105.e9.

Archambault, J., and Friesen, J.D. (1993). Genetics of Eukaryotic RNA Polymerases I, II, and III. MICROBIOL REV *57*, 22.

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of Cryptic Unstable Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. Mol. Cell *23*, 841–851.

Arudchandran, A., Cerritelli, S., Narimatsu, S., Itaya, M., Shin, D.Y., Shimada, Y., and Crouch, R.J. (2000). The absence of ribonuclease H1 or H2 alters the sensitivity of Saccharomyces cerevisiae to hydroxyurea, caffeine and ethyl methanesulphonate: implications for roles of RNases H in DNA replication and repair. Genes Cells Devoted Mol. Cell. Mech. *5*, 789–802.

Ayyagari, R., Gomes, X.V., Gordenin, D.A., and Burgers, P.M.J. (2003). Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 AND DNA2. J. Biol. Chem. *278*, 1618–1625.

Azvolinsky, A., Giresi, P.G., Lieb, J.D., and Zakian, V.A. (2009). Highly Transcribed RNA Polymerase II Genes Are Impediments to Replication Fork Progression in Saccharomyces cerevisiae. Mol. Cell *34*, 722–734.

Bae, S.H., Bae, K.H., Kim, J.A., and Seo, Y.S. (2001). RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. Nature *412*, 456–461.

Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., et al. (2017). Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. Mol. Cell *66*, 38-49.e6.

Baharoglu, Z., Lestini, R., Duigou, S., and Michel, B. (2010). RNA polymerase mutations that facilitate replication progression in the rep uvrD recF mutant lacking two accessory replicative helicases. Mol. Microbiol. *77*, 324–336.

Balk, B., Dees, M., Bender, K., and Luke, B. (2014). The differential processing of telomeres in response to increased telomeric transcription and RNA-DNA hybrid accumulation. RNA Biol. *11*, 95–100.

Bambara, R.A., Murante, R.S., and Henricksen, L.A. (1997). Enzymes and reactions at the eukaryotic DNA replication fork. J. Biol. Chem. *272*, 4647–4650.

Baranello, L., Wojtowicz, D., Cui, K., Devaiah, B.N., Chung, H.-J., Chan-Salis, K.Y., Guha, R., Wilson, K., Zhang, X., Zhang, H., et al. (2016). RNA Polymerase II Regulates Topoisomerase 1 Activity to Favor Efficient Transcription. Cell *165*, 357–371.

Barillà, D., Lee, B.A., and Proudfoot, N.J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. *98*, 445–450.

Baxter, J., and Diffley, J.F.X. (2008). Topoisomerase II Inactivation Prevents the Completion of DNA Replication in Budding Yeast. Mol. Cell *30*, 790–802.

Becherel, O.J., Yeo, A.J., Stellati, A., Heng, E.Y.H., Luff, J., Suraweera, A.M., Woods, R., Fleming, J., Carrie, D., McKinney, K., et al. (2013). Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. PLoS Genet. *9*, e1003435.

Bedinger, P., Hochstrasser, M., Jongeneel, C.V., and Alberts, B.M. (1983). Properties of the T4 bacteriophage DNA replication apparatus: the T4 dda DNA helicase is required to pass a bound RNA polymerase molecule. Cell *34*, 115–123.

Bell, S.P., and Kaguni, J.M. (2013). Helicase Loading at Chromosomal Origins of Replication. Cold Spring Harb. Perspect. Biol. *5*, a010124.

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

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128–134.

Belotserkovskii, B.P., Liu, R., Tornaletti, S., Krasilnikova, M.M., Mirkin, S.M., and Hanawalt, P.C. (2010). Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. Proc. Natl. Acad. Sci. U. S. A. *107*, 12816–12821.

Beranek, D.T. (1990). Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. Mutat. Res. *231*, 11–30.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., and Lemaitre, J.-M. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat. Struct. Mol. Biol. *19*, 837–844.

Bianchi, V., Pontis, E., and Reichard, P. (1986). Changes of deoxyribonucleoside triphosphate pools induced by hydroxyurea and their relation to DNA synthesis. J. Biol. Chem. *261*, 16037–16042.

Bienroth, S., Keller, W., and Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. EMBO J. *12*, 585–594.

Boguslawski, S.J., Smith, D.E., Michalak, M.A., Mickelson, K.E., Yehle, C.O., Patterson, W.L., and Carrico, R.J. (1986). Characterization of monoclonal antibody to DNA.RNA and its application to immunodetection of hybrids. J. Immunol. Methods *89*, 123–130.

Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Géli, V., de Almeida, S.F., and Palancade, B. (2017). Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability. Mol. Cell *67*, 608-621.e6.

Botchan, P., Wang, J.C., and Echols, H. (1973). Effect of circularity and superhelicity on transcription from bacteriophagelambda DNA. Proc. Natl. Acad. Sci. U. S. A. *70*, 3077–3081.

Brambati, A., Zardoni, L., Achar, Y.J., Piccini, D., Galanti, L., Colosio, A., Foiani, M., and Liberi, G. (2018). Dormant origins and fork protection mechanisms rescue sister forks arrested by transcription. Nucleic Acids Res. *46*, 1227–1239.

Branzei, D., and Foiani, M. (2007). Interplay of replication checkpoints and repair proteins at stalled replication forks. DNA Repair *6*, 994–1003.

Breker, M., Gymrek, M., and Schuldiner, M. (2013). A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. J. Cell Biol. *200*, 839–850.

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

Brewer, B.J., and Fangman, W.L. (1988). A replication fork barrier at the 3′ end of yeast ribosomal RNA genes. Cell *55*, 637–643.

Brewer, B.J., Lockshon, D., and Fangman, W.L. (1992). The arrest of replication forks in the rDNA of yeast occurs independently of transcription. Cell *71*, 267–276.

Broach, J.R., Li, Y.Y., Feldman, J., Jayaram, M., Abraham, J., Nasmyth, K.A., and Hicks, J.B. (1983). Localization and sequence analysis of yeast origins of DNA replication. Cold Spring Harb. Symp. Quant. Biol. *47 Pt 2*, 1165–1173.

Brown, P.O., and Cozzarelli, N.R. (1981). Catenation and knotting of duplex DNA by type 1 topoisomerases: a mechanistic parallel with type 2 topoisomerases. Proc. Natl. Acad. Sci. U. S. A. *78*, 843–847.

Buchman, A.R., Kimmerly, W.J., Rine, J., and Kornberg, R.D. (1988). Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in Saccharomyces cerevisiae. Mol. Cell. Biol. *8*, 210–225.

Buratowski, S. (2003). The CTD code. Nat. Struct. Biol. *10*, 679–680.

Cadoret, J.-C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.-N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc. Natl. Acad. Sci. U. S. A. *105*, 15837–15842.

Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A., and Labib, K. (2005). Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. Genes Dev. *19*, 1905–1919.

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

Caron, P.R. (1999). Compendium of DNA topoisomerase sequences. Methods Mol. Biol. Clifton NJ *94*, 279–316.

Carrasco-Salas, Y., Malapert, A., Sulthana, S., Molcrette, B., Chazot-Franguiadakis, L., Bernard, P., Chédin, F., Faivre-Moskalenko, C., and Vanoosthuyse, V. (2019). The extruded non-template strand determines the architecture of R-loops. Nucleic Acids Res. gkz341.

Carroll, K.L., Pradhan, D.A., Granek, J.A., Clarke, N.D., and Corden, J.L. (2004). Identification of cis Elements Directing Termination of Yeast Nonpolyadenylated snoRNA Transcripts. Mol. Cell. Biol. *24*, 6241–6252.

Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. RNA *13*, 361–373.

Castel, S.E., Ren, J., Bhattacharjee, S., Chang, A.-Y., Sánchez, M., Valbuena, A., Antequera, F., and Martienssen, R.A. (2014). Dicer Promotes Transcription Termination at Sites of Replication Stress to Maintain Genome Stability. Cell *159*, 572–583.

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

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., et al. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res. *21*, 1438–1449.

Cayrou, C., Ballester, B., Peiffer, I., Fenouil, R., Coulombe, P., Andrau, J.-C., Helden, J. van, and Méchali, M. (2015). The chromatin environment shapes DNA replication origin organization and defines origin classes. Genome Res. *25*, 1873–1885.

Cerritelli, S.M., and Crouch, R.J. (1995). The non-RNase H domain of Saccharomyces cerevisiae RNase H1 binds double-stranded RNA: magnesium modulates the switch between double-stranded RNA binding and RNase H activity. RNA N. Y. N *1*, 246–259.

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

Cerritelli, S.M., Frolova, E.G., Feng, C., Grinberg, A., Love, P.E., and Crouch, R.J. (2003). Failure to produce mitochondrial DNA results in embryonic lethality in Rnaseh1 null mice. Mol. Cell *11*, 807–815.

Chabes, A., Georgieva, B., Domkin, V., Zhao, X., Rothstein, R., and Thelander, L. (2003). Survival of DNA damage in yeast directly depends on increased dNTP levels allowed by relaxed feedback inhibition of ribonucleotide reductase. Cell *112*, 391–401.

Challal, D., Barucco, M., Kubik, S., Feuerbach, F., Candelli, T., Geoffroy, H., Benaksas, C., Shore, D., and Libri, D. (2018). General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. Mol. Cell *72*, 955-969.e7.

Champoux, J.J., and Dulbecco, R. (1972). An activity from mammalian cells that untwists superhelical DNA--a possible swivel for DNA replication (polyoma-ethidium bromide-mouse-embryo cells-dye binding assay). Proc. Natl. Acad. Sci. U. S. A. *69*, 143–146.

Chan, Y.A., Aristizabal, M.J., Lu, P.Y.T., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014). Genome-Wide Profiling of Yeast DNA:RNA Hybrid Prone Sites with DRIP-Chip. PLOS Genet. *10*, e1004288.

Chang, F., Theis, J.F., Miller, J., Nieduszynski, C.A., Newlon, C.S., and Weinreich, M. (2008). Analysis of Chromosome III Replicators Reveals an Unusual Structure for the ARS318 Silencer Origin and a Conserved WTW Sequence within the Origin Recognition Complex Binding Site. Mol. Cell. Biol. *28*, 5071–5081.

Changela, A., DiGate, R.J., and Mondragón, A. (2007). Structural studies of E. coli topoisomerase III-DNA complexes reveal a novel type IA topoisomerase-DNA conformational intermediate. J. Mol. Biol. *368*, 105–118.

Chastain, P.D., Makhov, A.M., Nossal, N.G., and Griffith, J. (2003). Architecture of the replication complex and DNA loops at the fork generated by the bacteriophage t4 proteins. J. Biol. Chem. *278*, 21276–21285.

Chatterji, M., Unniraman, S., Maxwell, A., and Nagaraja, V. (2000). The additional 165 amino acids in the B protein of Escherichia coli DNA gyrase have an important role in DNA binding. J. Biol. Chem. *275*, 22888–22894.

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

Chédin, F., Hartono, S.R., Sanz, L.A., and Vanoosthuyse, V. (2021). Best practices for the visualization, mapping, and manipulation of R-loops. EMBO J. *40*.

Chen, H.-T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat. Struct. Mol. Biol. *14*, 696–703.

Chen, L., Chen, J.-Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Mol. Cell *68*, 745-757.e5.

Chen, Y.-Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., et al. (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). Am. J. Hum. Genet. *74*, 1128–1135.

Cheng, B., Zhu, C.-X., Ji, C., Ahumada, A., and Tse-Dinh, Y.-C. (2003). Direct Interaction between Escherichia coli RNA Polymerase and the Zinc Ribbon Domains of DNA Topoisomerase I*. J. Biol. Chem. *278*, 30705–30710.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, D700-705.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: The exosome. Biochim. Biophys. Acta BBA - Gene Regul. Mech. *1829*, 552–560.

Cho, J.-E., and Jinks-Robertson, S. (2017). Ribonucleotides and Transcription-Associated Mutagenesis in Yeast. J. Mol. Biol. *429*, 3156–3167.

Chon, H., Vassilev, A., DePamphilis, M.L., Zhao, Y., Zhang, J., Burgers, P.M., Crouch, R.J., and Cerritelli, S.M. (2009). Contributions of the two accessory subunits, RNASEH2B and RNASEH2C, to the activity and properties of the human RNase H2 complex. Nucleic Acids Res. *37*, 96–110.

Chong, J.P.J., Thömmes, P., and Blow, J.J. (1996). The role of MCM/P1 proteins in the licensing of DNA replication. Trends Biochem. Sci. *21*, 102–106.

Ciccia, A., and Elledge, S.J. (2010). The DNA Damage Response: Making It Safe to Play with Knives. Mol. Cell *40*, 179–204.

Clarke, D.J., Segal, M., Jensen, S., and Reed, S.I. (2001). Mec1p regulates Pds1p levels in S phase: complex coordination of DNA replication and mitosis. Nat. Cell Biol. *3*, 619–627.

Cobb, J.A., Bjergbaek, L., Shimada, K., Frei, C., and Gasser, S.M. (2003). DNA polymerase stabilization at stalled replication forks requires Mec1 and the RecQ helicase Sgs1. EMBO J. *22*, 4325–4336.

Cobb, J.A., Schleker, T., Rojas, V., Bjergbaek, L., Tercero, J.A., and Gasser, S.M. (2005). Replisome instability, fork collapse, and gross chromosomal rearrangements arise synergistically from Mec1 kinase and RecQ helicase mutations. Genes Dev. *19*, 3055–3069.

Cohen, S., Puget, N., Lin, Y.-L., Clouaire, T., Aguirrebengoa, M., Rocher, V., Pasero, P., Canitrot, Y., and Legube, G. (2018). Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. Nat. Commun. *9*, 533.

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

Conover, H.N., Lujan, S.A., Chapman, M.J., Cornelio, D.A., Sharif, R., Williams, J.S., Clark, A.B., Camilo, F., Kunkel, T.A., and Argueso, J.L. (2015). Stimulation of Chromosomal Rearrangements by Ribonucleotides. Genetics *201*, 951–961.

Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. (2000). A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics *154*, 557–571.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. Science *322*, 1845–1848.

Costantino, L., and Koshland, D. (2018). Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. Mol. Cell *71*, 487-497.e3.

Coster, G., and Diffley, J.F.X. (2017). Bidirectional eukaryotic DNA replication is established by quasi-symmetrical helicase loading. Science *357*, 314–318.

Coster, G., Frigola, J., Beuron, F., Morris, E.P., and Diffley, J.F.X. (2014). Origin licensing requires ATP binding and hydrolysis by the MCM replicative helicase. Mol. Cell *55*, 666–677.

Crabbé, L., Thomas, A., Pantesco, V., De Vos, J., Pasero, P., and Lengronne, A. (2010). Analysis of replication profiles reveals key role of RFC-Ctf18 in yeast replication stress response. Nat. Struct. Mol. Biol. *17*, 1391–1397.

Crisona, N.J., Strick, T.R., Bensimon, D., Croquette, V., and Cozzarelli, N.R. (2000). Preferential relaxation of positively supercoiled DNA by E. coli topoisomerase IV in single-molecule and ensemble measurements. Genes Dev. *14*, 2881–2892.

Crow, Y.J., Leitch, A., Hayward, B.E., Garner, A., Parmar, R., Griffith, E., Ali, M., Semple, C., Aicardi, J., Babul-Hirji, R., et al. (2006). Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. Nat. Genet. *38*, 910–916.

Darst, S.A. (2001). Bacterial RNA polymerase. Curr. Opin. Struct. Biol. *11*, 155–162.

Davis, C.A., and Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomycescerevisiae. Proc. Natl. Acad. Sci. *103*, 3262–3267.

Deibler, R.W., Rahmati, S., and Zechiedrich, E.L. (2001). Topoisomerase IV, alone, unknots DNA in E. coli. Genes Dev. *15*, 748–761.

Delan-Forino, C., Schneider, C., and Tollervey, D. (2017). Transcriptome-wide analysis of alternative routes for RNA substrates into the exosome complex. PLoS Genet. *13*, e1006699.

Delgado, S., Gómez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J. *17*, 2426–2435.

DeMarini, D.J., Winey, M., Ursic, D., Webb, F., and Culbertson, M.R. (1992). SEN1, a positive effector of tRNA-splicing endonuclease in Saccharomyces cerevisiae. Mol. Cell. Biol. *12*, 2154–2164.

Dengl, S., and Cramer, P. (2009). Torpedo Nuclease Rat1 Is Insufficient to Terminate RNA Polymerase II in Vitro*. J. Biol. Chem. *284*, 21270–21279.

Deshpande, A.M., and Newlon, C.S. (1996). DNA Replication Fork Pause Sites Dependent on Transcription. Science *272*, 1030–1033.

Dewar, J.M., and Walter, J.C. (2017). Mechanisms of DNA replication termination. Nat. Rev. Mol. Cell Biol. *18*, 507–516.

Dhar, M.K., Sehgal, S., and Kaul, S. (2012). Structure, replication efficiency and fragility of yeast ARS elements. Res. Microbiol. *163*, 243–253.

Diffley, J.F.X., and Cocker, J.H. (1992). Protein-DNA interactions at a yeast replication origin. Nature *357*, 169–172.

Diffley, J.F.X., Cocker, J.H., Dowell, S.J., and Rowley, A. (1994). Two steps in the assembly of complexes at yeast replication origins in vivo. Cell *78*, 303–316.

DiGate, R.J., and Marians, K.J. (1989). Molecular cloning and DNA sequence analysis of Escherichia coli topB, the gene encoding topoisomerase III*. J. Biol. Chem. *264*, 17924–17930.

DiGate, R.J., and Marians, K.J. (1992). Escherichia coli topoisomerase III-catalyzed cleavage of RNA. J. Biol. Chem. *267*, 20532–20535.

van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., et al. (2011). XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature *475*, 114–117.

Dimude, J.U., Midgley-Smith, S.L., Stein, M., and Rudolph, C.J. (2016). Replication Termination: Containing Fork Fusion-Mediated Pathologies in Escherichia coli. Genes *7*, E40.

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

Duggin, I.G., Wake, R.G., Bell, S.D., and Hill, T.M. (2008). The replication fork trap and termination of chromosome replication. Mol. Microbiol. *70*, 1323–1333.

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev. *18*, 1618–1629.

Duquette, M.L., Huber, M.D., and Maizels, N. (2007). G-rich proto-oncogenes are targeted for genomic instability in B-cell lymphomas. Cancer Res. *67*, 2586–2594.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in E. coli. Cell *146*, 533–543.

Dvir, A. (2002). Promoter escape by RNA polymerase II. Biochim. Biophys. Acta 16.

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev. *32*, 127–139.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

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

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. Trends Genet. TIG *24*, 280–288.

Eichner, J., Chen, H.-T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. EMBO J. *29*, 706–716.

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

El Hage, A., Webb, S., Kerr, A., and Tollervey, D. (2014). Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. PLoS Genet. *10*, e1004716.

Ellwood, M., and Nomura, M. (1982). Chromosomal locations of the genes for rRNA in Escherichia coli K-12. J. Bacteriol. *149*, 458–468.

Evans, S.P., and Bycroft, M. (1999). NMR structure of the N-terminal domain of Saccharomyces cerevisiae RNase HI reveals a fold with a strong resemblance to the N-terminal domain of ribosomal protein L9. J. Mol. Biol. *291*, 661–669.

Fedoroff OYu, null, Salazar, M., and Reid, B.R. (1993). Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. J. Mol. Biol. *233*, 509–523.

Felipe-Abrio, I., Lafuente-Barquero, J., Garcia-Rubio, M.L., and Aguilera, A. (2015). RNA polymerase II contributes to preventing transcription-mediated replication fork stalls. EMBO J. *34*, 236–250.

Ferguson, B.M., and Fangman, W.L. (1992). A position effect on the time of replication origin activation in yeast. Cell *68*, 333–339.

Figiel, M., Chon, H., Cerritelli, S.M., Cybulska, M., Crouch, R.J., and Nowotny, M. (2011). The structural and biochemical characterization of human RNase H2 complex reveals the molecular basis for substrate recognition and Aicardi-Goutières syndrome defects. J. Biol. Chem. *286*, 10540–10550.

Finkel, J.S., Chinchilla, K., Ursic, D., and Culbertson, M.R. (2010). Sen1p Performs Two Genetically Separable Functions in Transcription and Processing of U5 Small Nuclear RNA in Saccharomyces cerevisiae. Genetics *184*, 107–118.

Foltman, M., Evrin, C., De Piccoli, G., Jones, R.C., Edmondson, R.D., Katou, Y., Nakato, R., Shirahige, K., and Labib, K. (2013). Eukaryotic replisome components cooperate to process histones during chromosome replication. Cell Rep. *3*, 892–904.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Mol. Cell *60*, 256–267.

French, S. (1992). Consequences of replication fork movement through transcription units in vivo. Science *258*, 1362–1365.

French, S.L., Sikes, M.L., Hontz, R.D., Osheim, Y.N., Lambert, T.E., El Hage, A., Smith, M.M., Tollervey, D., Smith, J.S., and Beyer, A.L. (2011). Distinguishing the roles of Topoisomerases I and II in relief of transcription-induced torsional stress in yeast rRNA genes. Mol. Cell. Biol. *31*, 482–494.

Funnell, B.E., Baker, T.A., and Kornberg, A. (1986). Complete enzymatic replication of plasmids containing the origin of the Escherichia coli chromosome. J. Biol. Chem. *261*, 5616–5624.

Ganem, C., Devaux, F., Torchet, C., Jacq, C., Quevillon-Cheruel, S., Labesse, G., Facca, C., and Faye, G. (2003). Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast. EMBO J. *22*, 1588–1598.

Gangloff, S., McDonald, J.P., Bendixen, C., Arthur, L., and Rothstein, R. (1994). The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. Mol. Cell. Biol. *14*, 8391–8398.

Gangloff, S., de Massy, B., Arthur, L., Rothstein, R., and Fabre, F. (1999). The essential role of yeast topoisomerase III in meiosis depends on recombination. EMBO J. *18*, 1701–1711.

Gao, F., and Zhang, C.-T. (2007). DoriC: a database of oriC regions in bacterial genomes. Bioinforma. Oxf. Engl. *23*, 1866–1867.

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

García-Rodríguez, L.J., De Piccoli, G., Marchesi, V., Jones, R.C., Edmondson, R.D., and Labib, K. (2015). A conserved Polε binding module in Ctf18-RFC is required for S-phase checkpoint activation downstream of Mec1. Nucleic Acids Res. *43*, 8830–8838.

García-Rubio, M.L., and Aguilera, A. (2012). Topological constraints impair RNA polymerase II transcription and causes instability of plasmid-borne convergent genes. Nucleic Acids Res. *40*, 1050–1064.

Gellert, M., Mizuuchi, K., O'Dea, M.H., and Nash, H.A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. Proc. Natl. Acad. Sci. U. S. A. *73*, 3872–3876.

Gellert, M., Fisher, L.M., and O'Dea, M.H. (1979). DNA gyrase: purification and catalytic properties of a fragment of gyrase B protein. Proc. Natl. Acad. Sci. U. S. A. *76*, 6289–6293.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. Mol. Cell *45*, 814–825.

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

Gloor, J.W., Balakrishnan, L., and Bambara, R.A. (2010). Flap endonuclease 1 mechanism analysis indicates flap base binding prior to threading. J. Biol. Chem. *285*, 34922–34931.

Glover, T.W., Berger, C., Coyle, J., and Echo, B. (1984). DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. Hum. Genet. *67*, 136–142.

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

Goodrich, J.A., and Tjian, R. (1994). Transcription Factors IIE and IIH and ATP Hydrolysis Direct Promoter Clearance by RNA Poiymeke II. 12.

Goto, T., and Wang, J.C. (1982). Yeast DNA topoisomerase II. An ATP-dependent type II topoisomerase that catalyzes the catenation, decatenation, unknotting, and relaxation of double-stranded DNA rings. J. Biol. Chem. *257*, 5866–5872.

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

Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause Sites Promote Transcriptional Termination of Mammalian RNA Polymerase II. Mol. Cell. Biol. *26*, 3986–3996.

Groth, P., Ausländer, S., Majumder, M.M., Schultz, N., Johansson, F., Petermann, E., and Helleday, T. (2010). Methylated DNA Causes a Physical Block to Replication Forks Independently of Damage Signalling, O6-Methylguanine or DNA Single-Strand Breaks and Results in DNA Damage. J. Mol. Biol. *402*, 70–82.

Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. Mol. Cell *48*, 409–421.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat. Struct. Mol. Biol. *11*, 394–403.

Hamdan, S.M., Loparo, J.J., Takahashi, M., Richardson, C.C., and van Oijen, A.M. (2009). Dynamics of DNA replication loops reveal temporal control of lagging-strand synthesis. Nature *457*, 336–339.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell *170*, 774-786.e19.

Hampsey, M. (1998). Molecular Genetics of the RNA Polymerase II General Transcriptional Machinery. Microbiol. Mol. Biol. Rev. *62*, 465–503.

Han, Z., Jasnovidova, O., Haidara, N., Tudek, A., Kubicek, K., Libri, D., Stefl, R., and Porrua, O. (2020). Termination of non-coding transcription in yeast relies on both an RNA Pol II CTD interaction domain and a CTD-mimicking region in Sen1. EMBO J.

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nat. Rev. Mol. Cell Biol. *18*, 263–273.

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K.M., Eaton, M.L., et al. (2015). BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair. Mol. Cell *57*, 636–647.

Hausen, P., and Stein, H. (1970). Ribonuclease H. An enzyme degrading the RNA moiety of DNA-RNA hybrids. Eur. J. Biochem. *14*, 278–283.

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

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

Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? Genes Dev. *7*, 1291–1308.

Hiasa, H., and Marians, K.J. (1996). Two distinct modes of strand unlinking during theta-type DNA replication. J. Biol. Chem. *271*, 21529–21535.

Ho, B., Baryshnikova, A., and Brown, G.W. (2018). Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. Cell Syst. *6*, 192-205.e3.

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., and Stefl, R. (2011). Recognition of Transcription Termination Signal by the Nuclear Polyadenylated RNA-binding (NAB) 3 Protein*. J. Biol. Chem. *286*, 3645–3657.

Holm, C., Goto, T., Wang, J.C., and Botstein, D. (1985). DNA topoisomerase II is required at the time of mitosis in yeast. Cell *41*, 553–563.

Holstege, F.C.P., Fiedler, U., and Timmers, H.T.M. (1997). Three transitions in the RNA polymerase II transcription complex during initiation. EMBO J. *16*, 7468–7480.

Hong, X., Cadwell, G.W., and Kogoma, T. (1995). Escherichia coli RecG and RecA proteins in R-loop formation. EMBO J. *14*, 2385–2392.

Horikoshi, M., Bertuccioli, C., Takada, R., Wang, J., Yamamoto, T., and Roeder, R.G. (1992). Transcription factor TFIID induces DNA bending upon binding to the TATA element. Proc. Natl. Acad. Sci. *89*, 1060–1064.

Horowitz, D.S., and Wang, J.C. (1987). Mapping the active site tyrosine of Escherichia coli DNA gyrase. J. Biol. Chem. *262*, 5339–5344.

Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., and Tollervey, D. (2006). Microarray detection of novel nuclear RNA substrates for the exosome. Yeast Chichester Engl. *23*, 439–454.

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

Hu, F., Wang, Y., Liu, D., Li, Y., Qin, J., and Elledge, S.J. (2001). Regulation of the Bub2/Bfa1 GAP complex by Cdc5 and cell cycle checkpoints. Cell *107*, 655–665.

Hu, Y., Lu, X., Zhou, G., Barnes, E.L., and Luo, G. (2009). Recql5 plays an important role in DNA replication and cell survival after camptothecin treatment. Mol. Biol. Cell *20*, 114–123.

Huang, R.Y., and Kowalski, D. (1993). A DNA unwinding element and an ARS consensus comprise a replication origin within a yeast chromosome. EMBO J. *12*, 4521–4531.

Huang, F.-T., Yu, K., Balter, B.B., Selsing, E., Oruc, Z., Khamlichi, A.A., Hsieh, C.-L., and Lieber, M.R. (2007). Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. Mol. Cell. Biol. *27*, 5921–5932.

Huertas, P., and Aguilera, A. (2003). Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. Mol. Cell *12*, 711–721.

Hunter, G.O., Fox, M.J., Smith-Kinnaman, W.R., Gogol, M., Fleharty, B., and Mosley, A.L. (2016). Phosphatase Rtr1 Regulates Global Levels of Serine 5 RNA Polymerase II C-Terminal Domain Phosphorylation and Cotranscriptional Histone Methylation. Mol. Cell. Biol. *36*, 2236–2245.

Huppert, J.L., and Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res. *35*, 406–413.

Huppert, J.L., Bugaut, A., Kumari, S., and Balasubramanian, S. (2008). G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res. *36*, 6260–6268.

Itoh, T., and Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc. Natl. Acad. Sci. *77*, 2450–2454.

Ivessa, A.S., Zhou, J.Q., and 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*, 479–489.

Ivessa, A.S., Lenzmeier, B.A., Bessler, J.B., Goudsouzian, L.K., Schnakenberg, S.L., and Zakian, V.A. (2003). The Saccharomyces cerevisiae Helicase Rrm3p Facilitates Replication Past Nonhistone Protein-DNA Complexes. Mol. Cell *12*, 1525–1536.

Jacob, F., Brenner, S., and Cuzin, F. (1963). On the Regulation of DNA Replication in Bacteria. Cold Spring Harb. Symp. Quant. Biol. *28*, 329–348.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with Pervasive Transcription. Mol. Cell *52*, 473–484.

Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. Nat. Struct. Mol. Biol. *21*, 449–455.

Jeronimo, C., Bataille, A.R., and Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. Chem. Rev. *113*, 8491–8522.

Jia, H., Wang, X., Anderson, J.T., and Jankowsky, E. (2012). RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. Proc. Natl. Acad. Sci. *109*, 7292–7297.

Jongeneel, C.V., Formosa, T., Munn, M., and Alberts, B.M. (1984). Enzymological studies of the T4 replication proteins. Adv. Exp. Med. Biol. *179*, 17–33.

Joshi, R.S., Piña, B., and Roca, J. (2012). Topoisomerase II is required for the production of long Pol II gene transcripts in yeast. Nucleic Acids Res. *40*, 7907–7915.

Kamada, K. (2012). The GINS complex: structure and function. Subcell. Biochem. *62*, 135–156.

Kang, S., Warner, M.D., and Bell, S.P. (2014). Multiple Functions for Mcm2-7 ATPase Motifs During Replication Initiation. Mol. Cell *55*, 655–665.

Katayama, T., Kubota, T., Kurokawa, K., Crooke, E., and Sekimizu, K. (1998). The Initiator Function of DnaA Protein Is Negatively Regulated by the Sliding Clamp of the E. coli Chromosomal Replicase. Cell *94*, 61–71.

Katayama, T., Ozaki, S., Keyamura, K., and Fujimitsu, K. (2010). Regulation of the replication cycle: conserved and diverse regulatory systems for DnaA and oriC. Nat. Rev. Microbiol. *8*, 163–170.

Kato, J., Nishimura, Y., Imamura, R., Niki, H., Hiraga, S., and Suzuki, H. (1990). New topoisomerase essential for chromosome segregation in E. coli. Cell *63*, 393–404.

Katou, Y., Kanoh, Y., Bando, M., Noguchi, H., Tanaka, H., Ashikari, T., Sugimoto, K., and Shirahige, K. (2003). S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. Nature *424*, 1078–1083.

Keszthelyi, A., Minchell, N.E., and Baxter, J. (2016). The Causes and Consequences of Topological Stress during DNA Replication. Genes *7*, 134.

Kim, J., and Iyer, V.R. (2004). Global Role of TATA Box-Binding Protein Recruitment to Promoters in Mediating Gene Expression Profiles. Mol. Cell. Biol. *24*, 8104–8112.

Kim, R.A., and Wang, J.C. (1992). Identification of the yeast TOP3 gene product as a single strand-specific DNA topoisomerase. J. Biol. Chem. *267*, 17178–17185.

Kim, J.H., Lee, B.B., Oh, Y.M., Zhu, C., Steinmetz, L.M., Lee, Y., Kim, W.K., Lee, S.B., Buratowski, S., and Kim, T. (2016). Modulation of mRNA and lncRNA expression dynamics by the Set2–Rpd3S pathway. Nat. Commun. *7*, 13534.

Kim, M., Ahn, S.-H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004a). Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. EMBO J. *23*, 354–364.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004b). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522.

Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct Pathways for snoRNA and mRNA Termination. Mol. Cell *24*, 723–734.

Kim, T., Xu, Z., Clauder-Münster, S., Steinmetz, L.M., and Buratowski, S. (2012). Set3 HDAC Mediates Effects of Overlapping Noncoding Transcription on Gene Induction Kinetics. Cell *150*, 1158–1169.

Kobayashi, T., and Horiuchi, T. (1996). **A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities**. Genes Cells *1*, 465–474.

Kobayashi, T., Hidaka, M., Nishizawa, M., and Horiuchi, T. (1992). Identification of a site required for DNA replication fork blocking activity in the rRNA gene cluster in Saccharomyces cerevisiae. Mol. Gen. Genet. MGG *233*, 355–362.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes Dev. *14*, 2452–2460.

Komata, M., Bando, M., Araki, H., and Shirahige, K. (2009). The direct binding of Mrc1, a checkpoint mediator, to Mcm6, a replication helicase, is essential for the replication checkpoint against methyl methanesulfonate-induced stress. Mol. Cell. Biol. *29*, 5008–5019.

Kowalski, D., and Eddy, M.J. (1989). The DNA unwinding element: a novel, cis-acting component that facilitates opening of the Escherichia coli replication origin. EMBO J. *8*, 4335–4344.

Kreuzer, K.N., and Brister, J.R. (2010). Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. Virol. J. *7*, 358.

Kreuzer, K.N., and Cozzarelli, N.R. (1979). Escherichia coli mutants thermosensitive for deoxyribonucleic acid gyrase subunit A: effects on deoxyribonucleic acid replication, transcription, and bacteriophage growth. J. Bacteriol. *140*, 424–435.

Kreuzer, K.N., and Cozzarelli, N.R. (1980). Formation and resolution of DNA catenanes by DNA gyrase. Cell *20*, 245–254.

Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes Dev. *26*, 1891–1896.

Kuehner, J.N., and Brow, D.A. (2008). Regulation of a Eukaryotic Gene by GTP-Dependent Start Site Selection and Transcription Attenuation. Mol. Cell *31*, 201–211.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. Nat. Rev. Mol. Cell Biol. *12*, 283–294.

Kuempel, P.L., Duerr, S.A., and Seeley, N.R. (1977). Terminus region of the chromosome in Escherichia coli inhibits replication forks. Proc. Natl. Acad. Sci. U. S. A. *74*, 3927–3931.

Kumar, C., Batra, S., Griffith, J.D., and Remus, D. (2021). The interplay of RNA:DNA hybrid structure and G-quadruplexes determines the outcome of R-loop-replisome collisions. ELife *10*, e72286.

Kyburz, A., Sadowski, M., Dichtl, B., and Keller, W. (2003). The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3′-end formation. Nucleic Acids Res. *31*, 3936–3945.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA Degradation by the Exosome Is Promoted by a Nuclear Polyadenylation Complex. Cell *121*, 713–724.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat. Rev. Mol. Cell Biol. *18*, 548–562.

Lai, F., Damle, S.S., Ling, K.K., and Rigo, F. (2020). Directed RNase H Cleavage of Nascent Transcripts Causes Transcription Termination. Mol. Cell *77*, 1032-1043.e4.

Lambert, S., Froget, B., and Carr, A. (2007). Arrested replication fork processing: Interplay between checkpoints and recombination. DNA Repair *6*, 1042–1061.

Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., and Merrikh, H. (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell *170*, 787-799.e18.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lascaris, R.F., Mager, W.H., and Planta, R.J. (1999). DNA-binding requirements of the yeast protein Rap1p as selected in silico from ribosomal protein gene promoter sequences. Bioinforma. Oxf. Engl. *15*, 267–277.

Le, T.T., Gao, X., Park, S. ha, Lee, J., Inman, J.T., Lee, J.H., Killian, J.L., Badman, R.P., Berger, J.M., and Wang, M.D. (2019). Synergistic Coordination of Chromatin Torsional Mechanics and Topoisomerase Activity. Cell *179*, 619-631.e15.

Lee, S.H., Kwong, A.D., Pan, Z.Q., and Hurwitz, J. (1991). Studies on the activator 1 protein complex, an accessory factor for proliferating cell nuclear antigen-dependent DNA polymerase delta. J. Biol. Chem. *266*, 594–602.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G.P., Itoh, T., Watanabe, Y., Shirahige, K., and Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature *430*, 573–578.

Leonaitė, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O., and Conti, E. (2017). Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. EMBO J. *36*, 1590–1604.

Letessier, A., Millot, G.A., Koundrioukoff, S., Lachagès, A.-M., Vogt, N., Hansen, R.S., Malfoy, B., Brison, O., and Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. Nature *470*, 120–123.

Li, H., and Stillman, B. (2012). The origin recognition complex: a biochemical and structural view. Subcell. Biochem. *62*, 37–58.

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

Li, X., and Manley, J.L. (2006). Cotranscriptional processes and their influence on genome stability. Genes Dev. *20*, 1838–1847.

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

Li, M., Pokharel, S., Wang, J.-T., Xu, X., and Liu, Y. (2015a). RECQ5-dependent SUMOylation of DNA topoisomerase I prevents transcription-associated genome instability. Nat. Commun. *6*, 6720.

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.-K., and Gao, N. (2015b). Structure of the eukaryotic MCM complex at 3.8 Å. Nature *524*, 186–191.

Li, X., Li, J., Harrington, J., Lieber, M.R., and Burgers, P.M. (1995). Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. J. Biol. Chem. *270*, 22109–22112.

Li, Z., Mondragón, A., Hiasa, H., Marians, K.J., and DiGate, R.J. (2000). Identification of a unique domain essential for Escherichia coli DNA topoisomerase III-catalysed decatenation of replication intermediates. Mol. Microbiol. *35*, 888–895.

Lidschreiber, M., Easter, A.D., Battaglia, S., Rodríguez-Molina, J.B., Casañal, A., Carminati, M., Baejen, C., Grzechnik, P., Maier, K.C., Cramer, P., et al. (2018). The APT complex is involved in non-coding RNA transcription and is distinct from CPF. Nucleic Acids Res. *46*, 11528–11538.

Lima, C.D., Wang, J.C., and Mondragón, A. (1994). Three-dimensional structure of the 67K N-terminal fragment of E. coli DNA topoisomerase I. Nature *367*, 138–146.

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

Lipford, J.R., and Bell, S.P. (2001). Nucleosomes Positioned by ORC Facilitate the Initiation of DNA Replication. Mol. Cell *7*, 21–30.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. U. S. A. *84*, 7024–7027.

Liu, L.F., Liu, C.-C., and Alberts, B.M. (1980). Type II DNA topoisomerases: Enzymes that can unknot a topologically knotted DNA molecule via a reversible double-strand break. Cell *19*, 697–707.

Lockhart, A., Pires, V.B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., Ulrich, H.D., and Luke, B. (2019). RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. Cell Rep. *29*, 2890-2900.e5.

Logan, J., Falck-Pedersen, E., Darnell, J.E., and Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. Proc. Natl. Acad. Sci. *84*, 8306–8310.

Longtine, M.I., Demarini, S., Wach, B., and Philippsen, P. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. 9.

Lopes, M., Cotta-Ramusino, C., Pellicioli, A., Liberi, G., Plevani, P., Muzi-Falconi, M., Newlon, C.S., and Foiani, M. (2001). The DNA replication checkpoint response stabilizes stalled replication forks. Nature *412*, 557–561.

López, V., Martínez-Robles, M.-L., Hernández, P., Krimer, D.B., and Schvartzman, J.B. (2012). Topo IV is the topoisomerase that knots and unknots sister duplexes during DNA replication. Nucleic Acids Res. *40*, 3563–3573.

Lopez-Mosqueda, J., Maas, N.L., Jonsson, Z.O., Defazio-Eli, L.G., Wohlschlegel, J., and Toczyski, D.P. (2010). Damage-induced phosphorylation of Sld3 is important to block late origin firing. Nature *467*, 479–483.

Lou, H., Komata, M., Katou, Y., Guan, Z., Reis, C.C., Budd, M., Shirahige, K., and Campbell, J.L. (2008). Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. Mol. Cell *32*, 106–117.

Louarn, J., Patte, J., and Louarn, J.M. (1977). Evidence for a fixed termination site of chromosome replication in Escherichia coli K12. J. Mol. Biol. *115*, 295–314.

Loya, T.J., O'Rourke, T.W., and Reines, D. (2012). A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3. Nucleic Acids Res. *40*, 7476–7491.

Loya, T.J., O'Rourke, T.W., and Reines, D. (2013). Yeast Nab3 Protein Contains a Self-assembly Domain Found in Human Heterogeneous Nuclear Ribonucleoprotein-C (hnRNP-C) That Is Necessary for Transcription Termination*. J. Biol. Chem. *288*, 2111–2117.

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

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription

termination factors with the RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. *17*, 1195–1201.

Luo, W., Johnson, A.W., and Bentley, D.L. (2006). The role of Rat1 in coupling mRNA 3′-end processing to transcription termination: implications for a unified allosteric–torpedo model. Genes Dev. *20*, 954–965.

Luse, D.S., and Jacob, G.A. (1987). Abortive initiation by RNA polymerase II in vitro at the adenovirus 2 major late promoter. J. Biol. Chem. *262*, 14990–14997.

Lynch, P.J., Fraser, H.B., Sevastopoulos, E., Rine, J., and Rusche, L.N. (2005). Sum1p, the origin recognition complex, and the spreading of a promoter-specific repressor in Saccharomyces cerevisiae. Mol. Cell. Biol. *25*, 5920–5932.

Lynn, R., Giaever, G., Swanberg, S.L., and Wang, J.C. (1986). Tandem regions of yeast DNA topoisomerase II share homology with different subunits of bacterial gyrase. Science *233*, 647–649.

Mackiewicz, P., Zakrzewska-Czerwińska, J., Zawilak, A., Dudek, M.R., and Cebrat, S. (2004). Where does bacterial replication start? Rules for predicting the oriC region. Nucleic Acids Res. *32*, 3781–3791.

Magdalou, I., Lopez, B.S., Pasero, P., and Lambert, S.A.E. (2014). The causes of replication stress and their consequences on genome stability and cell fate. Semin. Cell Dev. Biol. *30*, 154–164.

Mailand, N., Gibbs-Seymour, I., and Bekker-Jensen, S. (2013). Regulation of PCNA–protein interactions for genome stability. Nat. Rev. Mol. Cell Biol. *14*, 269–282.

Majka, J., and Burgers, P.M.J. (2003). Yeast Rad17/Mec3/Ddc1: a sliding clamp for the DNA damage checkpoint. Proc. Natl. Acad. Sci. U. S. A. *100*, 2249–2254.

Majka, J., Niedziela-Majka, A., and Burgers, P.M.J. (2006). The checkpoint clamp activates Mec1 kinase during initiation of the DNA damage checkpoint. Mol. Cell *24*, 891–901.

Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., and Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. ELife *4*, e06722.

Malig, M., Hartono, S.R., Giafaglione, J.M., Sanz, L.A., and Chedin, F. (2020). Ultra-deep Coverage Single-molecule R-loop Footprinting Reveals Principles of R-loop Formation. J. Mol. Biol. *432*, 2271–2288.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3′-end-processing endonuclease. Nature *444*, 953–956.

Marians, K.J. (1987). DNA gyrase-catalyzed decatenation of multiply linked DNA dimers. J. Biol. Chem. *262*, 10362–10368.

Marquardt, S., Hazelbaker, D.Z., and Buratowski, S. (2011). Distinct RNA degradation pathways and 3′ extensions of yeast non-coding RNA species. Transcription *2*, 145–154.

Marszalek, J., and Kaguni, J.M. (1994). DnaA protein directs the binding of DnaB protein in initiation of DNA replication in Escherichia coli. J. Biol. Chem. *269*, 4883–4890.

McGarry, K.C., Ryan, V.T., Grimwade, J.E., and Leonard, A.C. (2004). Two discriminatory binding sites in the Escherichia coli replication origin are required for DNA strand opening by initiator DnaA-ATP. Proc. Natl. Acad. Sci. *101*, 2811–2816.

McGuffee, S.R., Smith, D.J., and Whitehouse, I. (2013). Quantitative, Genome-Wide Analysis of Eukaryotic Replication Initiation and Termination. Mol. Cell *50*, 123–135.

Meinhart, A., and Cramer, P. (2004). Recognition of RNA polymerase II carboxy-terminal domain by 3′-RNA-processing factors. Nature *430*, 223–226.

Merrikh, C.N., Brewer, B.J., and Merrikh, H. (2015). The B. subtilis Accessory Helicase PcrA Facilitates DNA Replication through Transcription Units. PLoS Genet. *11*, e1005289.

Meryet-Figuiere, M., Alaei-Mahabadi, B., Ali, M.M., Mitra, S., Subhash, S., Pandey, G.K., Larsson, E., and Kanduri, C. (2014). Temporal separation of replication and transcription during S-phase progression. Cell Cycle Georget. Tex *13*, 3241–3248.

Mesner, L.D., Valsakumar, V., Karnani, N., Dutta, A., Hamlin, J.L., and Bekiranov, S. (2011). Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. Genome Res. *21*, 377–389.

Mimura, S., Komata, M., Kishi, T., Shirahige, K., and Kamura, T. (2009). SCFDia2 regulates DNA replication forks during S-phase in budding yeast. EMBO J. *28*, 3693–3705.

Mirkin, E.V., and Mirkin, S.M. (2007). Replication Fork Stalling at Natural Impediments. Microbiol. Mol. Biol. Rev. *71*, 13–35.

Mirkin, E.V., Castro Roa, D., Nudler, E., and Mirkin, S.M. (2006). Transcription regulatory elements are punctuation marks for DNA replication. Proc. Natl. Acad. Sci. *103*, 7276–7281.

Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 Helicase Protects the Genome from Transcription-Associated Instability. Mol. Cell *41*, 21–32.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The Exosome: A Conserved Eukaryotic RNA Processing Complex Containing Multiple 3′→5′ Exoribonucleases. Cell *91*, 457–466.

Mohanty, B.K., Bairwa, N.K., and Bastia, D. (2006). The Tof1p-Csm3p protein complex counteracts the Rrm3p helicase to control replication termination of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. *103*, 897–902.

Mondragón, A., and DiGate, R. (1999). The structure of Escherichia coli DNA topoisomerase III. Structure *7*, 1373–1383.

Morafraile, E.C., Diffley, J.F.X., Tercero, J.A., and Segurado, M. (2015). Checkpoint-dependent RNR induction promotes fork restart after replicative stress. Sci. Rep. *5*, 7886.

Moreira, M.-C., Klur, S., Watanabe, M., Németh, A.H., Le Ber, I., Moniz, J.-C., Tranchant, C., Aubourg, P., Tazir, M., Schöls, L., et al. (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. Nat. Genet. *36*, 225–227.

Moreno, S.P., Bailey, R., Campion, N., Herron, S., and Gambus, A. (2014). Polyubiquitylation drives replisome disassembly at the termination of DNA replication. Science *346*, 477–481.

Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and Washburn, M.P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Mol. Cell *34*, 168–178.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Curr. Opin. Struct. Biol. *13*, 31–39.

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

Nag, A., Narsinh, K., and Martinson, H.G. (2007). The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. Nat. Struct. Mol. Biol. *14*, 662–669.

Navadgi-Patil, V.M., and Burgers, P.M. (2009). The unstructured C-terminal tail of the 9-1-1 clamp subunit Ddc1 activates Mec1/ATR via two distinct mechanisms. Mol. Cell *36*, 743–753.

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 Phosphatase Subunit of the Cleavage and Polyadenylation Factor Is Essential for Transcription Termination on snoRNA Genes. Mol. Cell *29*, 577–587.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038–1042.

Nevers, A., Doyen, A., Malabat, C., Néron, B., Kergrohen, T., Jacquier, A., and Badis, G. (2018). Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. Nucleic Acids Res. *46*, 6009–6025.

Nguyen, H.D., Yadav, T., Giri, S., Saez, B., Graubert, T.A., and Zou, L. (2017). Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. Mol. Cell *65*, 832-847.e4.

Nick McElhinny, S.A., Gordenin, D.A., Stith, C.M., Burgers, P.M.J., and Kunkel, T.A. (2008). Division of Labor at the Eukaryotic Replication Fork. Mol. Cell *30*, 137–144.

Nievera, C., Torgue, J.J.-C., Grimwade, J.E., and Leonard, A.C. (2006). SeqA Blocking of DnaA-oriC Interactions Ensures Staged Assembly of the E. coli Pre-RC. Mol. Cell *24*, 581–592.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G., and Burley, S.K. (1995). Crystal structure of a TFIIB–TBP–TATA-element ternary complex. Nature *377*, 119–128.

Nitiss, J.L. (1998). Investigating the biological functions of DNA topoisomerases in eukaryotic cells. Biochim. Biophys. Acta BBA - Gene Struct. Expr. *1400*, 63–81.

Nomura, M., Morgan, E.A., and Jaskunas, S.R. (1977). Genetics of Bacterial Ribosomes. Annu. Rev. Genet. *11*, 297–347.

Nowotny, M., Gaidamakov, S.A., Crouch, R.J., and Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell *121*, 1005–1016.

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

O'Connell, K., Jinks-Robertson, S., and Petes, T.D. (2015). Elevated Genome-Wide Instability in Yeast Mutants Lacking RNase H Activity. Genetics *201*, 963–975.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., and Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. Proc. Natl. Acad. Sci. U. S. A. *59*, 598–605.

Orozco, I.J., Kim, S.J., and Martinson, H.G. (2002). The Poly(A) Signal, without the Assistance of Any Downstream Element, Directs RNA Polymerase II to Pause in Vivo and Then to Release Stochastically from the Template*. J. Biol. Chem. *277*, 42899–42911.

Osheim, Y.N., Proudfoot, N.J., and Beyer, A.L. (1999). EM Visualization of Transcription by RNA Polymerase II: Downstream Termination Requires a Poly(A) Signal but Not Transcript Cleavage. Mol. Cell 9.

Osheim, Y.N., Sikes, M.L., and Beyer, A.L. (2002). EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3′ end cleavage of nascent transcripts. Chromosoma *111*, 1–12.

Ozeri-Galai, E., Lebofsky, R., Rahat, A., Bester, A.C., Bensimon, A., and Kerem, B. (2011). Failure of origin activation in response to fork stalling leads to chromosomal instability at fragile sites. Mol. Cell *43*, 122–131.

Paeschke, K., Capra, J.A., and Zakian, V.A. (2011). DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. Cell *145*, 678–691.

Pancevac, C., Goldstone, D.C., Ramos, A., and Taylor, I.A. (2010). Structure of the Rna15 RRM-RNA complex reveals the molecular basis of GU specificity in transcriptional 3′-end processing factors. Nucleic Acids Res. *38*, 3119–3132.

Pardo, B., Crabbé, L., and Pasero, P. (2016). Signaling Pathways of Replication Stress in Yeast. FEMS Yeast Res. fow101.

Parvin, J.D., and Sharp, P.A. (1993). DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. Cell *73*, 533–540.

Pato, M.L. (1975). Alterations of the rate of movement of deoxyribonucleic acid replication forks. J. Bacteriol. *123*, 272–277.

Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E., and Merrikh, H. (2013). Accelerated gene evolution through replication–transcription conflicts. Nature *495*, 512–515.

Paulsen, R.D., Soni, D.V., Wollman, R., Hahn, A.T., Yee, M.-C., Guan, A., Hesley, J.A., Miller, S.C., Cromwell, E.F., Solow-Cordero, D.E., et al. (2009). A Genome-wide siRNA Screen Reveals Diverse Cellular Processes and Pathways that Mediate Genome Stability. Mol. Cell *35*, 228–239.

Pavlov, Y.I., Frahm, C., Nick McElhinny, S.A., Niimi, A., Suzuki, M., and Kunkel, T.A. (2006). Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta. Curr. Biol. CB *16*, 202–207.

Pearson, E.L., and Moore, C.L. (2013). Dismantling Promoter-driven RNA Polymerase II Transcription Complexes in Vitro by the Termination Factor Rat1*. J. Biol. Chem. *288*, 19750–19759.

Pellegrini, L. (2012). The Pol α-primase complex. Subcell. Biochem. *62*, 157–169.

Pellicioli, A., and Foiani, M. (2005). Signal Transduction: How Rad53 Kinase Is Activated. Curr. Biol. *15*, R769–R771.

Peng, H., and Marians, K.J. (1993a). Escherichia coli topoisomerase IV. Purification, characterization, subunit structure, and subunit interactions. J. Biol. Chem. *268*, 24481–24490.

Peng, H., and Marians, K.J. (1993b). Decatenation activity of topoisomerase IV during oriC and pBR322 DNA replication in vitro. Proc. Natl. Acad. Sci. U. S. A. *90*, 8571–8575.

Peng, H., and Marians, K.J. (1995). The interaction of Escherichia coli topoisomerase IV with DNA. J. Biol. Chem. *270*, 25286–25290.

Perera, R.L., Torella, R., Klinge, S., Kilkenny, M.L., Maman, J.D., and Pellegrini, L. (2013). Mechanism for priming DNA synthesis by yeast DNA polymerase α. ELife *2*, e00482.

Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. Nature *418*, 99–103.

Petes, T.D. (1979). Yeast ribosomal DNA genes are located on chromosome XII. Proc. Natl. Acad. Sci. U. S. A. *76*, 410–414.

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

Poli, J., Tsaponina, O., Crabbé, L., Keszthelyi, A., Pantesco, V., Chabes, A., Lengronne, A., and Pasero, P. (2012). dNTP pools determine fork progression and origin usage under replication stress. EMBO J. *31*, 883–894.

Pomerantz, R.T., and O'Donnell, M. (2010). What happens when replication and transcription complexes collide? Cell Cycle *9*, 2537–2543.

Popuri, V., Tadokoro, T., Croteau, D.L., and Bohr, V.A. (2013). Human RECQL5: guarding the crossroads of DNA replication and transcription and providing backup capability. Crit. Rev. Biochem. Mol. Biol. *48*, 289–299.

Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nat. Struct. Mol. Biol. *20*, 884–891.

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

Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R., and Libri, D. (2012). *In vivo* SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination: Novel determinants of degradative termination. EMBO J. *31*, 3935–3948.

Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends Genet. *32*, 508–522.

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

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

Promonet, A., Padioleau, I., Liu, Y., Sanz, L., Biernacka, A., Schmitz, A.-L., Skrzypczak, M., Sarrazin, A., Mettling, C., Rowicka, M., et al. (2020). Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. Nat. Commun. *11*, 3940.

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

Pursell, Z.F., Isoz, I., Lundström, E.-B., Johansson, E., and Kunkel, T.A. (2007). Yeast DNA Polymerase ε Participates in Leading-Strand DNA Replication. Science *317*, 127–130.

Qiu, H., Hu, C., and Hinnebusch, A.G. (2009). Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. Mol. Cell *33*, 752–762.

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication Dynamics of the Yeast Genome. Science *294*, 115–121.

Ranish, J.A., Yudkovsky, N., and Hahn, S. (1999). Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a postrecruitment role for the TATA box and TFIIB. Genes Dev. *13*, 49–63.

Rao, H., and Stillman, B. (1995). The origin recognition complex interacts with a bipartite DNA binding site within yeast replicators. Proc. Natl. Acad. Sci. *92*, 2224–2228.

Rawal, C.C., Zardoni, L., Di Terlizzi, M., Galati, E., Brambati, A., Lazzaro, F., Liberi, G., and Pellicioli, A. (2020). Senataxin Ortholog Sen1 Limits DNA:RNA Hybrid Accumulation at DNA Double-Strand Breaks to Control End Resection and Repair Fidelity. Cell Rep. *31*, 107603.

Reaban, M.E., Lebowitz, J., and Griffin, J.A. (1994). Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. J. Biol. Chem. *269*, 21850–21857.

Reijns, M.A.M., Bubeck, D., Gibson, L.C.D., Graham, S.C., Baillie, G.S., Jones, E.Y., and Jackson, A.P. (2011). The structure of the human RNase H2 complex defines key interaction interfaces relevant to enzyme function and human disease. J. Biol. Chem. *286*, 10530–10539.

Remus, D., and Diffley, J.F. (2009). Eukaryotic DNA replication control: Lock and load, then fire. Curr. Opin. Cell Biol. *21*, 771–777.

Rhind, N., and Gilbert, D.M. (2013). DNA Replication Timing. Cold Spring Harb. Perspect. Biol. *5*, a010132.

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

Roda, R.H., Rinaldi, C., Singh, R., Schindler, A.B., and Blackstone, C. (2014). Ataxia with oculomotor apraxia type 2 fibroblasts exhibit increased susceptibility to oxidative DNA damage. J. Clin. Neurosci. Off. J. Neurosurg. Soc. Australas. *21*, 1627–1631.

Rouse, J., and Jackson, S.P. (2002). Lcd1p Recruits Mec1p to DNA Lesions In Vitro and In Vivo. Mol. Cell *9*, 857–869.

Rowley, A., Cocker, J.H., Harwood, J., and Diffley, J.F. (1995). Initiation complex assembly at budding yeast replication origins begins with the recognition of a bipartite sequence by limiting amounts of the initiator, ORC. EMBO J. *14*, 2631–2641.

Ryan, K., Calvo, O., and Manley, J.L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3′ processing endonuclease. RNA *10*, 565–573.

Rychlik, M.P., Chon, H., Cerritelli, S.M., Klimek, P., Crouch, R.J., and Nowotny, M. (2010). Crystal structures of RNase H2 in complex with nucleic acid reveal the mechanism of RNA-DNA junction recognition and cleavage. Mol. Cell *40*, 658–670.

Sabouri, N., McDonald, K.R., Webb, C.J., Cristea, I.M., and Zakian, V.A. (2012). DNA replication through hard-to-replicate sites, including both highly transcribed RNA Pol II and Pol III genes, requires the S. pombe Pfh1 helicase. Genes Dev. *26*, 581–593.

Saffer, L.D., and Miller, O.L. (1986). Electron microscopic study of Saccharomyces cerevisiae rDNA chromatin replication. Mol. Cell. Biol. *6*, 1148–1157.

Salazar, M., Fedoroff, O.Y., Miller, J.M., Ribeiro, N.S., and Reid, B.R. (1993). The DNA strand in DNA.RNA hybrid duplexes is neither B-form nor A-form in solution. Biochemistry *32*, 4207–4215.

Salisbury, J., Hutchison, K.W., and Graber, J.H. (2006). A multispecies comparison of the metazoan 3′-processing downstream elements and the CstF-64 RNA recognition motif. BMC Genomics *7*, 55.

Sanchez, Y., Desany, B.A., Jones, W.J., Liu, Q., Wang, B., and Elledge, S.J. (1996). Regulation of RAD53 by the ATM-like kinases MEC1 and TEL1 in yeast cell cycle checkpoint pathways. Science *271*, 357–360.

Santos-Pereira, J.M., Herrero, A.B., García-Rubio, M.L., Marín, A., Moreno, S., and Aguilera, A. (2013). The Npl3 hnRNP prevents R-loop-mediated transcription–replication conflicts and genome instability. Genes Dev. *27*, 2445–2458.

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

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

Sawitzke, J.A., and Austin, S. (2000). Suppression of chromosome segregation defects of Escherichia coli muk mutants by mutations in topoisomerase I. Proc. Natl. Acad. Sci. U. S. A. *97*, 1671–1676.

Schaper, S., and Messer, W. (1995). Interaction of the Initiator Protein DnaA of Escherichia coli with Its DNA Target *. J. Biol. Chem. *270*, 17622–17626.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-Wide Mapping of Yeast RNA Polymerase II Termination. PLoS Genet. *10*, e1004632.

Schroeder, S.C., Schwer, B., Shuman, S., and Bentley, D. (2000). Dynamic association of capping enzymes with transcribing RNA polymerase II. Genes Dev. *14*, 2435–2440.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. Cell *155*, 1075–1087.

Sequeira-Mendes, J., and Gómez, M. (2012). On the opportunistic nature of transcription and replication initiation in the metazoan genome. BioEssays News Rev. Mol. Cell. Dev. Biol. *34*, 119–125.

Sequeira-Mendes, J., Díaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gómez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet. *5*, e1000446.

Shaban, N.M., Harvey, S., Perrino, F.W., and Hollis, T. (2010). The structure of the mammalian RNase H2 complex provides insight into RNA.NA hybrid processing to prevent immune dysfunction. J. Biol. Chem. *285*, 3617–3624.

Shaw, N.N., and Arya, D.P. (2008). Recognition of the unique structure of DNA:RNA hybrids. Biochimie *90*, 1026–1039.

Shcherbakova, P.V., and Pavlov, Y.I. (1996). 3′ -> 5′ Exonucleases of DNA Polymerases ε and δ Correct Base Analog Induced DNA Replication Errors on opposite DNA Strands in Saccharomyces Cerevisiae. Genetics *142*, 717–726.

Shearwin, K.E., Callen, B.P., and Egan, J.B. (2005). Transcriptional interference – a crash course. Trends Genet. TIG *21*, 339–345.

Shyian, M., Albert, B., Zupan, A.M., Ivanitsa, V., Charbonnet, G., Dilg, D., and Shore, D. (2020). Fork pausing complex engages topoisomerases at the replisome. Genes Dev. *34*, 87–98.

Simon, A.C., Zhou, J.C., Perera, R.L., van Deursen, F., Evrin, C., Ivanova, M.E., Kilkenny, M.L., Renault, L., Kjaer, S., Matak-Vinković, D., et al. (2014). A Ctf4 trimer couples the CMG helicase to DNA polymerase α in the eukaryotic replisome. Nature *510*, 293–297.

Siow, C.C., Nieduszynska, S.R., Müller, C.A., and Nieduszynski, C.A. (2012). OriDB, the DNA replication origin database updated and extended. Nucleic Acids Res. *40*, D682-686.

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

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. Mol. Cell *42*, 794–805.

Skryabin, K.G., Eldarov, M.A., Larionov, V.L., Bayev, A.A., Klootwijk, J., de Regt, V.C., Veldman, G.M., Planta, R.J., Georgiev, O.I., and Hadjiolov, A.A. (1984). Structure and function of the nontranscribed spacer regions of yeast rDNA. Nucleic Acids Res. *12*, 2955–2968.

Smirnov, E., Borkovec, J., Kováčik, L., Svidenská, S., Schröfel, A., Skalníková, M., Švindrych, Z., Křížek, P., Ovesný, M., Hagen, G.M., et al. (2014). Separation of replication and transcription domains in nucleoli. J. Struct. Biol. *188*, 259–266.

Sogo, J.M., Lopes, M., and Foiani, M. (2002). Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. Science *297*, 599–602.

Sollier, J., Stork, C.T., García-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. Mol. Cell *56*, 777–785.

Soren, B.C., Dasari, J.B., Ottaviani, A., Iacovelli, F., and Fiorani, P. (2020). Topoisomerase IB: a relaxing enzyme for stressed DNA. Cancer Drug Resist. *3*, 18–25.

Speck, C., Chen, Z., Li, H., and Stillman, B. (2005). ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. Nat. Struct. Mol. Biol. *12*, 965–971.

Srivatsan, A., Tehranchi, A., MacAlpine, D.M., and Wang, J.D. (2010). Co-orientation of replication and transcription preserves genome integrity. PLoS Genet. *6*, e1000810.

Stein, H., and Hausen, P. (1969). Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. Science *166*, 393–395.

Steinmetz, E.J., and Brow, D.A. (1998). Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. Proc. Natl. Acad. Sci. *95*, 6699–6704.

Stewart, L., Redinbo, M.R., Qiu, X., Hol, W.G., and Champoux, J.J. (1998). A model for the mechanism of human topoisomerase I. Science *279*, 1534–1541.

Stinchcomb, D.T., Struhl, K., and Davis, R.W. (1979). Isolation and characterisation of a yeast chromosomal replicator. Nature *282*, 39–43.

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

Stivers, J.T., Harris, T.K., and Mildvan, A.S. (1997). Vaccinia DNA topoisomerase I: evidence supporting a free rotation mechanism for DNA supercoil relaxation. Biochemistry *36*, 5212–5222.

Struhl, K. (1995). YEAST TRANSCRIPTIONAL REGULATORY MECHANISMS. 25.

Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. Biochemistry *34*, 11211–11216.

Sugino, A., Higgins, N.P., Brown, P.O., Peebles, C.L., and Cozzarelli, N.R. (1978). Energy coupling in DNA gyrase and the mechanism of action of novobiocin. Proc. Natl. Acad. Sci. U. S. A. *75*, 4838–4842.

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. Science *340*, 619–621.

Suski, C., and Marians, K.J. (2008). Resolution of converging replication forks by RecQ and topoisomerase III. Mol. Cell *30*, 779–789.

Sutton, M.D., Carr, K.M., Vicente, M., and Kaguni, J.M. (1998). Escherichia coli DnaA Protein: THE N-TERMINAL DOMAIN AND LOADING OF DnaB HELICASE AT THEE. COLI CHROMOSOMAL ORIGIN*. J. Biol. Chem. *273*, 34255–34262.

Szyjka, S.J., Viggiani, C.J., and Aparicio, O.M. (2005). Mrc1 is required for normal progression of replication forks throughout chromatin in S. cerevisiae. Mol. Cell *19*, 691–697.

Tadokoro, T., and Kanaya, S. (2009). Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. FEBS J. *276*, 1482–1493.

Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A., and Araki, H. (2003). GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. Genes Dev. *17*, 1153–1165.

Tan, K., Zhou, Q., Cheng, B., Zhang, Z., Joachimiak, A., and Tse-Dinh, Y.-C. (2015). Structural basis for suppression of hypernegative DNA supercoiling by E. coli topoisomerase I. Nucleic Acids Res. *43*, 11031–11046.

Tanaka, K., and Russell, P. (2004). Cds1 phosphorylation by Rad3-Rad26 kinase is mediated by forkhead-associated domain interaction with Mrc1. J. Biol. Chem. *279*, 32079–32086.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature *445*, 328–332.

Tan-Wong, S.M., Dhir, S., and Proudfoot, N.J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. Mol. Cell *76*, 600-616.e6.

Tehranchi, A.K., Blankschien, M.D., Zhang, Y., Halliday, J.A., Srivatsan, A., Peng, J., Herman, C., and Wang, J.D. (2010). The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. Cell *141*, 595–605.

Tercero, J.A., and Diffley, J.F.X. (2001). Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. Nature *412*, 553–557.

Tercero, J.A., Labib, K., and Diffley, J.F. (2000). DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. EMBO J. *19*, 2082–2093.

Theis, J.F., and Newlon, C.S. (1997). The ARS309 chromosomal replicator of Saccharomyces cerevisiae depends on an exceptional ARS consensus sequence. Proc. Natl. Acad. Sci. *94*, 10786–10791.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription Termination and Nuclear Degradation of Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. Mol. Cell *23*, 853–864.

Thiebaut, M., Colin, J., Neil, H., Jacquier, A., Séraphin, B., Lacroute, F., and Libri, D. (2008). Futile Cycle of Transcription Initiation and Termination Modulates the Response to Nucleotide Shortage in S. cerevisiae. Mol. Cell *31*, 671–682.

Tourrière, H., and Pasero, P. (2007). Maintenance of fork integrity at damaged DNA and natural pause sites. DNA Repair *6*, 900–913.

Tse, Y., and Wang, J.C. (1980). E. coli and M. luteus DNA topoisomerase I can catalyze catenation or decatenation of double-stranded DNA rings. Cell *22*, 269–276.

Tsurimoto, T., and Stillman, B. (1990). Functions of replication factor C and proliferating-cell nuclear antigen: functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. Proc. Natl. Acad. Sci. *87*, 1023–1027.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., et al. (2014). Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. Mol. Cell *55*, 467–481.

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

Uemura, T., Ohkura, H., Adachi, Y., Morino, K., Shiozaki, K., and Yanagida, M. (1987). DNA topoisomerase II is required for condensation and separation of mitotic chromosomes in S. pombe. Cell *50*, 917–925.

Ursic, D., Chinchilla, K., Finkel, J.S., and Culbertson, M.R. (2004). Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. Nucleic Acids Res. *32*, 2441–2452.

Valentini, S.R., Weiss, V.H., and Silver, P.A. (1999). Arginine methylation and binding of Hrp1p to the efficiency element for mRNA 3′-end formation. RNA N. Y. N *5*, 272–280.

Vaňáčová, Š., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A New Yeast Poly(A) Polymerase Complex Involved in RNA Quality Control. PLOS Biol. *3*, e189.

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

Villa, F., Simon, A.C., Ortiz Bazan, M.A., Kilkenny, M.L., Wirthensohn, D., Wightman, M., Matak-Vinkovíc, D., Pellegrini, L., and Labib, K. (2016a). Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase. Mol. Cell *63*, 385–396.

Villa, M., Cassani, C., Gobbini, E., Bonetti, D., and Longhese, M.P. (2016b). Coupling end resection with the checkpoint response at DNA double-strand breaks. Cell. Mol. Life Sci. CMLS *73*, 3655–3663.

Villa, T., Barucco, M., Martin-Niclos, M.-J., Jacquier, A., and Libri, D. (2020). Degradation of Non-coding RNAs Promotes Recycling of Termination Factors at Sites of Transcription. Cell Rep. *32*, 107942.

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

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

Wahle, E. (1991). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. Cell *66*, 759–768.

Wang, J.C. (1971). Interaction between DNA and an Escherichia coli protein ω. J. Mol. Biol. *55*, 523-IN16.

Wang, J.C. (1996). DNA topoisomerases. Annu. Rev. Biochem. *65*, 635–692.

Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. Nat. Rev. Mol. Cell Biol. *3*, 430–440.

Wang, J.D., Berkmen, M.B., and Grossman, A.D. (2007). Genome-wide coorientation of replication and transcription reduces adverse effects on replication in Bacillus subtilis. Proc. Natl. Acad. Sci. U. S. A. *104*, 5608–5613.

Wang, S., Han, Z., Libri, D., Porrua, O., and Strick, T.R. (2019). Single-molecule characterization of extrinsic transcription termination by Sen1 helicase. Nat. Commun. *10*, 1545.

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

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5′ → 3′ exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522–525.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural Basis of Transcription: Nucleotide Selection by Rotation in the RNA Polymerase II Active Center. Cell *119*, 481–489.

Wigley, D.B., Davies, G.J., Dodson, E.J., Maxwell, A., and Dodson, G. (1991). Crystal structure of an N-terminal fragment of the DNA gyrase B protein. Nature *351*, 624–629.

Williams, J.S., Lujan, S.A., and Kunkel, T.A. (2016). Processing ribonucleotides incorporated during eukaryotic DNA replication. Nat. Rev. Mol. Cell Biol. *17*, 350–363.

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. Mol. Cell *54*, 601–612.

Wu, H., Lima, W.F., and Crooke, S.T. (2001). Investigating the structure of human RNase H1 by site-directed mutagenesis. J. Biol. Chem. *276*, 23547–23553.

Wu, J., Phatnani, H.P., Hsieh, T.-S., and Greenleaf, A.L. (2010). The phosphoCTD-interacting domain of Topoisomerase I. Biochem. Biophys. Res. Commun. *397*, 117–119.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. Cell *121*, 725–737.

Xie, J., Aiello, U., Clement, Y., Haidara, N., Girbig, M., Schmitzova, J., Pena, V., Müller, C.W., Libri, D., and Porrua, O. (2021). An integrated model for termination of RNA polymerase III transcription.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., and Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. Gene *389*, 52–65.

Yeo, A.J., Becherel, O.J., Luff, J.E., Graham, M.E., Richard, D., and Lavin, M.F. (2015). Senataxin controls meiotic silencing through ATR activation and chromatin remodeling. Cell Discov. *1*, 15025.

Yeung, R., and Smith, D.J. (2020). Determinants of Replication-Fork Pausing at tRNA Genes in *Saccharomyces cerevisiae*. Genetics *214*, 825–838.

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

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., and Corden, J.L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. Proc. Natl. Acad. Sci. *93*, 6975–6980.

Zaratiegui, M., Castel, S.E., Irvine, D.V., Kloc, A., Ren, J., Li, F., de Castro, E., Marín, L., Chang, A.-Y., Goto, D., et al. (2011). RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. Nature *479*, 135–138.

Zechiedrich, E.L., and Cozzarelli, N.R. (1995). Roles of topoisomerase IV and DNA gyrase in DNA unlinking during replication in Escherichia coli. Genes Dev. *9*, 2859–2869.

Zegerman, P., and Diffley, J.F.X. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281–285.

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

Zhang, Z., and Gilmour, D.S. (2006). Pcf11 Is a Termination Factor in Drosophila that Dismantles the Elongation Complex by Bridging the CTD of RNA Polymerase II to the Nascent Transcript. Mol. Cell *21*, 65–74.

Zhang, H., Rigo, F., and Martinson, H.G. (2015). Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Mol. Cell *59*, 437–448.

Zhang, Z., Fu, J., and Gilmour, D.S. (2005). CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3′-end processing factor, Pcf11. Genes Dev. *19*, 1572–1580.

Zhao, D.Y., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F.W., Zhong, G., Liu, K., Li, W., Moffat, J., et al. (2016). SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. Nature *529*, 48–53.

Zhao, J., Bacolla, A., Wang, G., and Vasquez, K.M. (2010). Non-B DNA structure-induced genetic instability and evolution. Cell. Mol. Life Sci. CMLS *67*, 43–62.

Zhou, C., Elia, A.E.H., Naylor, M.L., Dephoure, N., Ballif, B.A., Goel, G., Xu, Q., Ng, A., Chou, D.M., Xavier, R.J., et al. (2016). Profiling DNA damage-induced phosphorylation in budding yeast reveals diverse signaling networks. Proc. Natl. Acad. Sci. U. S. A. *113*, E3667-3675.

Zimmer, A.D., and Koshland, D. (2016). Differential roles of the RNases H in preventing chromosome instability. Proc. Natl. Acad. Sci. *113*, 12220–12225.

Zimmermann, M., Murina, O., Reijns, M.A.M., Agathanggelou, A., Challis, R., Tarnauskaitė, Ž., Muir, M., Fluteau, A., Aregger, M., McEwan, A., et al. (2018). CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. Nature *559*, 285–289.

Zou, L., and Elledge, S.J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. Science *300*, 1542–1548.

Zou, L., and Stillman, B. (1998). Formation of a Preinitiation Complex by S-phase Cyclin CDK-Dependent Loading of Cdc45p onto Chromatin. Science *280*, 593–596.