UNIVERSIDADE DE LISBOA
FACULDADE DE MEDICINA DE LISBOA

Transcription dynamics prevents RNA-mediated genomic instability through SRPK2-dependent DDX23 phosphorylation

Sree Rama Chaitanya Sridhara

Orientadores: Prof. Doutor Sérgio Alexandre Fernandes de Almeida
Prof.ª Doutora Maria do Carmo Salazar Velez Roque da Fonseca

Tese especialmente elaborada para obtenção do grau de
Doutor em Ciências Biomédicas
Especialidade Biologia Celular e Molecular

2017

UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA DE LISBOA

Transcription dynamics prevents RNA-mediated genomic instability through SRPK2-dependent DDX23 phosphorylation

Sree Rama Chaitanya Sridhara

Orientadores: Prof. Doutor Sérgio Alexandre Fernandes de Almeida

Prof.ª Doutora Maria do Carmo Salazar Velez Roque da Fonseca

Tese especialmente elaborada para obtenção do grau de Doutor em Ciências Biomédicas especialidade Biologia Celular e Molecular

Júri:

Presidente: Professor Doutor José Luís Bliebernicht Ducla Soares, Professor Catedrático em regime de *tenure* e Vice-presidente do Conselho Cientifico da Faculdade de Medicina da Universidade de Lisboa.

Vogais:

- Doutor Nicholas Proudfoot, *Professor of Molecular Biology* da University of Oxford, Reino Unido;
- Doutor Lars Erwin Theodoor Jansen, Investigador Principal do Instituto Gulbenkian de Ciência;
- Doutor Luís Filipe Ferreira Moita, Investigador Principal do Instituto Gulbenkian de Ciência;
- Doutor Claus Maria Azzalin, Investigador, *Group Leader*, do do Instituto de Medicina Molecular, unidade de investigação associada à Faculdade de Medicina da Universidade de Lisboa.
- Doutor João António Augusto Ferreira, Professor Associado da Faculdade de Medicina da Universidade de Lisboa;
- Doutor Sérgio Alexandre Fernandes de Almeida, Professor Auxiliar da Faculdade de Medicina da Universidade de Lisboa; (*Orientador*).

Marie Skłodowska-Curie actions and Fundação para a Ciência e a Tecnologia

2017

“ A impressão desta tese foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de dia *2017.05.23”*

The ending is nearer than you think, and it is already written. All that we have left to choose is the correct moment to begin.

Alan Moore

The present thesis entitled '**Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation**' is divided into five major chapters with a short summary preceding it in both Portuguese and English. Each chapter is divided into smaller sections that explain an experiment or an idea. A general introduction to the topic of research has been dealt in **Chapter 1** with a special emphasis on 'R-loops and RNA processing factors'. The results obtained pertaining to this dissertation are presented as three individual chapters − **Chapter 2, 3** and **4**. The central idea of this project is to comprehend the molecular mechanisms employed by cells to sense, signal and resolve R-loops. Each of **Chapter 2, 3** and **4** will describe the data related to **SRPK2 – *signal transducer***, **DDX23 – *molecular effector*** and **RNA Pol II – *sensor*** for R-loops, respectively. **Extended data table 1** − SRPK1 and SRPK2 ChIP-seq data and **Extended data table 2** – DDX23 mutational analysis were attached in electronic format. A general scope of the work with implications in cancer has been discussed in **Chapter 5**. All the technical details and materials used related to this study has been elaborately mentioned at the end as **Appendix A** in order not to break the flow of the main text. Additionally, references, abbreviations and published articles were mentioned at the end from **Appendices B-D**. All the files pertaining to this thesis are attached in electronic format for further reference. This thesis is profusely filled with paraphrases and anecdotes that emphasize the meaning of a chapter in order to provide an organic reading.

Sree Rama Chaitanya

Thanks to my scientific mentor Prof. Sérgio de Almeida.

I can never teach you anything, you will have to teach yourself, but I can help perhaps in giving expression to that thought. **Vivekananda**

To learn by force or harshness; but direct them to it by what amuses their minds, so that you may be better able to discover with accuracy the peculiar bent of the genius of each. **Plato**

Thanks to my co-supervisor Prof. Dr. Maria Carmo-fonseca and her group members for their kind support. Thanks to my past and present lab mates for a collective and friendly environment. Hearty thanks to all my well wishers in iMM. Thanks to my thesis committee members Dr. João Ferreira and Dr. Luis Moita for their valuable inputs. Immense thanks to all the people who read my thesis and gave valuable suggestions. Special thanks to all my collaborators for sharing reagents and protocols.

Work of any sort is a collective effort, whether it is bones and flesh of our own or from others. **Anonymous**

As friendly as people could be from worlds as different as theirs and ours at that particular moment in our lives. **'Che' Guevara**

Thanks to funding agencies Marie Skłodowska-Curie actions and Fundação para a Ciência e a Tecnologia. Thanks to iMM for the support.

When there is wealth, a little effort accomplishes the task. **Chanakya**

Making money isn't hard in itself... What's hard is to earn it doing something worth devoting one's life to. **Carlos Ruiz Zafón**.

Thanks to Lisboa and Santarém :D.

As nações todas são mistérios.
Cada uma é todo o mundo a sós.
Fernando Pessoa

Thanks to my family for their love and support.

Love is an endless mystery, because there is no reasonable cause that could explain it. **Tagore**

In your love my salvation lies. **Alexi Murdoch**

This work could not have been done without the love and support from the best part of my life − Friends!

Depth of friendship does not depend on length of acquaintance. **Tagore**

There is self-interest behind every friendship. There is no friendship without self-interests. This is a bitter truth. **Chanakya**

friendship is a relationship that has no formal shape, there are no rules or obligations or bonds as in marriage or the family, it is held together by neither law nor property nor blood, there is no glue in it but mutual liking. It is therefore rare. **Wallace Stegner**

Resumo

Durante a transcrição, os transcritos de RNA são sintetizados pela RNA polimerase II (RNA Pol II) a partir da informação contida na cadeia de DNA. A síntese de RNA ocorre dentro da bolha de transcrição onde as duas cadeias de DNA são fisicamente separadas e o RNA nascente forma um híbrido RNA-DNA de aproximadamente 8 pares de bases com a cadeia molde. No entanto, a estrutura molecular da RNA Pol II assegura que o transcrito nascente seja fisicamente separado do DNA aquando da sua saída do local activo da enzima. Por esta razão, foi, até há pouco tempo, amplamente aceite que as estruturas híbridas de RNA-DNA seriam apenas produtos transientes da transcrição. Contudo, uma década de investigação demonstrou que híbridos de RNA-DNA se formam igualmente não é jusante da bolha de transcrição, originando, juntamente com cadeia de DNA desemparelhada, estruturas que se designam R-loops. Os R-loops são importantes para a regulação da dinâmica da transcrição e em diversos processos celulares, tais como a recombinação e a reparação de DNA. No entanto, os R-loops representam também uma grande ameaça à estabilidade genómica e devem por isso ser mantidos dentro de níveis fisiológicos.

Em condições normais, as células regulam os níveis fisiológicos de R-loops através de enzimas como helicases de DNA/RNA e nucleases (como RNase H) de forma a separar ou digerir a fração de RNA dos R-loops, respetivamente, com o intuito de manter a conformação do DNA nativo. Além disso, são conhecidos vários fatores de processamento de RNA que impedem a formação de R-loops, sequestrando o RNA nascente e impedindo-o de se ligar ao DNA, prevenindo assim a instabilidade genómica associada à formação de R-loops. Por exemplo, o fator de ligação SRSF1 é conhecido por prevenir a instabilidade genómica mediada por R-loops. Contudo, as cascatas de sinalização e as vias moleculares que conduzem à sua mobilização e ativação são essencialmente desconhecidas.

A atividade de diversos fatores de processamento de RNA é regulada através da fosforilação que é, em muitos casos, conduzida por proteínas quinase de serina/arginina (SRPK) 1 e 2. Estas quinases fosforilam domínios de arginina-serina (SR) de vários fatores de processamento de RNA classificados como proteínas-SR e regulam a sua função. O ciclo de fosforilação e desfosforilação prejudica a regulação da função das proteínas-SR e a sua ligação atempada aos elementos reguladores *cis* do RNA nascente. No entanto, a função das SRPK1 e SRPK2 no metabolismo de R-loops e na estabilidade do genoma ainda não foi desvendada.

Neste trabalho identificámos uma nova função associada à proteína SRPK2 na prevenção da instabilidade genómica. Para além disso, verificámos que a instabilidade genómica observada na ausência de SRPK2 é um fenómeno dependente da transcrição, mais concretamente da formação de R-loops. Na procura da base molecular desta observação, identificámos a DDX23, uma helicase DEAD box com um domínio RS na porção N-terminal, como sendo o substrato da proteína SRPK2 responsável por prevenir a instabilidade genómica mediada por R-loops. Mostrámos que a perda individual destas duas proteínas resulta na acumulação de níveis elevados de R-loops, que posteriormente conduzem a quebras no DNA. Adicionalmente, células com depleção de SRPK2 e DDX23 possuem aberrações cromossómicas condicentes com a elevada instabilidade genómica observada. O nosso trabalho revelou que a fosforilação da proteína DDX23 é necessária e suficiente para restaurar a estabilidade genómica em células com depleção de SRPK2 através de um processo que envolve a sua função de helicase de RNA. Embora a proteína DDX23 faça parte do complexo ribonucleoproteíco U5 snRNP do spliceossoma, demonstrámos que o seu papel na supressão de R-loops não necessita de um U5 snRNP funcional, uma vez que a depleção de PRP8 ou PRP6, componentes principais de U5 snRNP, não gera instabilidade genómica.

De forma a investigar o processo através do qual a SRPK2 reconhecia a presença dos R-loops, fosforilando a DDX23 de modo a suprimi-los, identificámos a pausa da RNA Pol II promovida pelos R-loops como factor central neste mecanismo. Assim, na presença de um R-loop, a RNA Pol II pausa a sua elongação, o que denuncia a localização do R-loop e promove o recrutamento de SRPK2 e a fosforilação da DDX23. Assim, este processo constitui um novo mecanismo zelador da integridade genómica que atua de forma a evitar danos no DNA que poderiam potenciar a formação de tumores. Suportando este modelo, observámos que a inexistência do locus do DDX23 é uma característica frequente no Carcinoma Adenóide Cístico, um cancro muito agressivo das glândulas salivares e com oportunidades de tratamento limitadas devido ao conhecimento incompleto da sua base genética e molecular. Em resumo, este trabalho permite-nos concluir que oscilações na dinâmica da RNA polimerase II constituem um sensor molecular de R-loops que pode prevenir eventos potencialmente catastróficos que inviabilizam a manutenção da estabilidade genómica.

Summary

During transcription, RNA is synthesized by RNA polymerase II (RNA Pol II) using the information contained in the template DNA strand. RNA synthesis occurs inside the transcription bubble where the two strands of DNA are physically separated and the nascent transcript is hybridized to template strand through a ~8 base pair (bp) RNA-DNA hybrid. Latter, the molecular structure of the elongating RNA Pol II ensures that the nascent transcript leaves the transcription bubble physically separated from the template DNA. Therefore, it was widely believed that RNA-DNA hybrids were only transient by-products of transcription. However, the past decade of research has shown those R-loops − RNA-DNA hybrids and a displaced single-stranded DNA−are abundant structures that play important roles in regulating gene expression and DNA recombination. Besides these physiological roles, R-loops also pose great threats to genome stability and should therefore be tightly regulated.

Cells regulate the level of R-loops by employing enzymes like DNA/RNA helicases and nucleases to resolve the RNA moiety of R-loops. Additionally, various RNA processing factors are also known to prevent R-loop formation by sequestering the nascent RNA, thus preventing it from re-annealing to the template DNA. However, the signalling cascades and molecular switches that lead to the mobilization and activation of R-loop suppressors are essentially unknown.

The activity of several RNA processing factors is regulated through phosphorylation, which in many cases is driven by serine/arginine protein kinases (SRPK) 1 and 2. These kinases phosphorylate the arginine-serine (RS) domain of various RNA processing factors classified as SR-proteins. Phosphorylation/de-phosphorylation cycles regulate the function of SR-proteins and the timely binding to *cis* regulatory elements on the nascent RNA. However, the roles of both SRPK1 and SRPK2 in R-loop metabolism and genome stability have not yet been investigated. Herein, we identify a new role for SRPK2 in preventing genomic instability through a mechanism involving the suppression of R-loops. We show that phosphorylation of the DEAD box helicase-DDX23 is necessary and sufficient to restore the genome stability in *SRPK2*-deficient cells through a process that requires its helicase activity.

DDX23 is part of the spliceosomal U5 small nuclear ribonucleoprotein complex (U5 snRNP). We found that the role of DDX23 in suppressing R-loops does not require a functional U5 snRNP as depletion of either *PRP8* or *PRP6*, core components of the U5

snRNP, does not drive genomic instability. Altogether, we found that phosphorylation of DDX23 by SRPK2 is a crucial event to maintain cellular R-loop levels, failure of which severely compromises genome integrity.

Co-transcriptional R-loops impact on RNA Pol II transcription dynamics. For instance, they are involved in promoter-proximal pausing and promote efficient transcription termination by slowing down RNA Pol II facilitating the timely recruitment of termination factors. Here we describe a new pathway that employs oscillations in RNA Pol II dynamics as a molecular sensor to signal the location of R-loops and nucleate SRPK2-dependent DDX23 phosphorylation. This may constitute a new genome caretaker mechanism that would operate to ward off against tumour-driving DNA damage. Supporting this view, we observed that loss of *DDX23* is a prevalent feature of adenoid cystic carcinoma, an aggressive salivary gland cancer with very limited treatment options mostly due to our incomplete understanding of its genomic foundations and molecular basis.

All these fifty years of conscious brooding have brought me no nearer to the answer to the question, 'What are light quanta?' Nowadays every Tom, Dick and Harry thinks he knows it, but he is mistaken!

Albert Einstein

I learned very early the difference between knowing the name of something and knowing something.

Richard Feynman

I just don't exactly feel I'm the master of my fate and the captain of my soul. Are you ready (to begin)?

Wallace Stegner

Chapter 1. Introduction

Highlights

- Genome instability is a hallmark of cancer.
- R-loops are a major endogenous source of genome instability.
- Defects in RNA processing machinery lead to unscheduled R-loop formation and genome instability.
- Serine/arginine protein kinases 1 and 2 (SRPK1 and SRPK2) modulate the function of RNA processing factors.

1.1 Genome instability is a hallmark of cancer

Our DNA is constantly bombarded by various physical or chemical insults causing potential damage (**Fig.1**) (Helleday et al., 2014). Like, ultraviolet (UV) rays from sunlight are an example of a physical genotoxic agent. UV triggers photo-reactions in DNA that cause lesions like photo-products and/or pyrimidine dimers. A gross estimate suggests that 10^5 UV photo-products are generated in keratinocytes exposed on a sunny day (Ciccia and Elledge, 2010). These lesions alter the DNA structure and interfere with DNA replication and eventually cause mutations in the genome (Cleaver et al., 2003).

DNA damage can also result from endogenous sources such as free radicals produced during normal cellular metabolism. Reactive oxygen and nitrogen species generated as metabolic by-products of general cell function can induce the formation of oxidative DNA products. Such products distort the DNA structure and threat genome stability eventually leading to spontaneous mutations and/or base loss in DNA (Ciccia and Elledge, 2010; Lindahl and Barnes, 2000).

Approximately 10^5 spontaneous DNA lesions can be experienced by every cell in one day (Hoeijmakers, 2015). Cells have to cope up with a constant threat to DNA and manage the damage that accumulates through time. However, the cumulative effect of various exogenous and endogenous sources of DNA damage challenges the genomic integrity beyond cells capability to handle such insults, leading to genome instability. Thus,

genomic instability can be defined as *an increased propensity for genomic alterations* (Shen, 2011; Aguilera and García-Muse, 2013) .

Genome surveillance mechanisms ensure that the genome integrity is kept on check by detecting and resolving defects in the DNA and controlling the rate of spontaneous mutations. These surveillance mechanisms monitor the genomic integrity and push damaged cells to either arrest their cell cycle for further repair or to apoptosis (Jackson and Bartek, 2009). The genome surveillance machinery consists of caretakers that help in: a) detecting damaged DNA and signaling it to the repair machinery, b) repair of damaged DNA c) hindering further damage (Ciccia and Elledge, 2010).

Cancer cells often augment the rate of mutations by increasing sensitivity to mutagenic agents concomitantly breaking down several components of the genome surveillance machinery (Jackson and Bartek, 2009; Liang et al., 2009). By doing so, cancer cells keep the cell cycle progression despite the presence of an unstable genome, thus acquiring mutations that in turn favor tumor progression (Shen, 2011). Accordingly, proteins involved in DNA repair pathways are mutated in several cancers leading to increased genome instability (Dietlein et al., 2014). In fact, abundant evidence suggests that genomic instability is instrumental for tumor progression, thus being classified as a ‘hallmark of cancer’ (Hanahan and Weinberg, 2011).

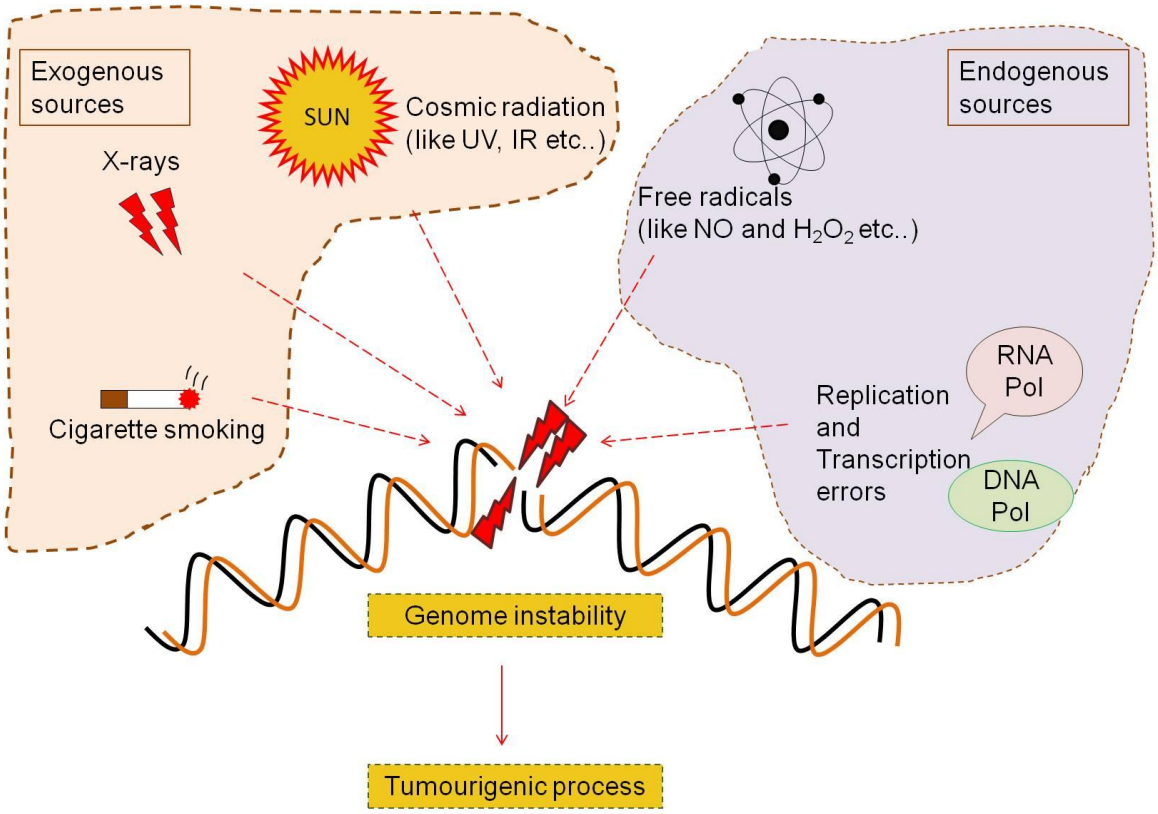


Figure 1. Genome instability is a hallmark of cancer. Various exogenous (e.g. X-rays from medical intervention, smoking and cosmic radiation) and endogenous sources (e.g. free radicals, activity of DNA polymerases (DNA Pol) and RNA Polymerases (RNA Pol)) constantly threat the integrity of our genome. Failure to repair damaged DNA (marked by a red thunder bolt) triggers tumourigenic process.

1.2 Transcription as an endogenous source of genome instability – a focus on R-loops

During the first step of gene expression, the DNA is transcribed by RNA polymerase II (RNA Pol II) to make an RNA molecule. Such transcription takes place on the same physical template as DNA replication, DNA repair and DNA recombination thereby physically and functionally connecting DNA and RNA metabolism. Cells segregate these different pathways in space and time, failure of which may cause collisions between transcription and replication machinery leading to genome instability (Aguilera, 2002). Transcription also poses threat to genome integrity by opening of DNA duplex and making it accessible to mutagenic agents thereby increasing the propensity for mutations (Aguilera and García-Muse, 2012). Recent evidence also suggests a direct role of RNA Pol II in contributing for different genome instability events including hyper-recombination, DNA damage sensitivity and/or DNA double-strand breaks (DSBs) (Felipe-Abrio et al., 2014).

Recent high-resolution cryo-electron microscopy of the transcribing RNA Pol II illustrates the interface between the nascent RNA and the template DNA within the polymerase where short RNA-DNA hybrids of ~8 base pairs (bp) are formed (**Fig.2a**) (Bernecky et al., 2016; Cheung and Cramer, 2012). Elongating RNA Pol II ensures that nascent RNA is physically separated from the transcribed DNA, preventing the formation of extended RNA-DNA hybrids. However, transcription creates negative supercoiling upstream RNA Pol II, which provides a window of opportunity for nascent RNA to hybridize with the template DNA. In this scenario, the resulting three-stranded nucleic acid structure − consisting of an RNA-DNA hybrid and a displaced single stranded DNA (ssDNA) – is termed R-loop (**Fig.2b**) (Hamperl and Cimprich, 2014).

Two models have been proposed for the formation of R-loops. The ‘extended RNA-DNA hybrid’ model, suggests that R-loops are basically an extension of the 8 bp RNA-DNA hybrid formed inside the elongating transcription bubble (Aguilera and García-Muse, 2012; Skourti-Stathaki and Proudfoot, 2014). However, structural data strongly suggest that the exit channel for the nascent RNA allows for physical separation between RNA and DNA (Bernecky et al., 2016; Westover et al., 2004), suggesting that R-loops cannot be formed as a mere extension of the 8 bp RNA-DNA hybrid. Although physically separated, RNA does stay in close proximity to the template DNA, which may facilitate the re-annealing of these two strands ‘outside’ the polymerase milieu, as proposed by the

alternative model for the formation of R-loops − the 'thread back' model. The 'thread back' model proposes that the nascent RNA invades the template DNA upstream the elongating RNA Pol II (Aguilera and García-Muse, 2012; Chédin, 2016; Skourti-Stathaki and Proudfoot, 2014).

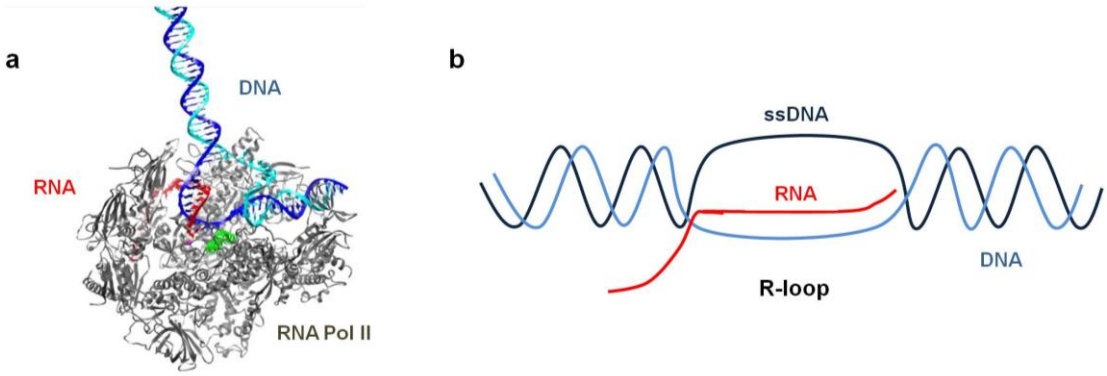


Figure 2. Schematic view of an R-loop structure. a) Ribbon diagram of an RNA-DNA hybrid within an elongating RNA Pol II (gray), with RNA (red) and DNA (shades of blue). Note that the exit channel of nascent RNA is different from that of the DNA. Adapted from Cheung and Cramer, 2012. **b)** Simple line diagram of an R-loop with RNA (red) annealed to the template strand of DNA (blue) and the displaced ssDNA (dark blue).

1.2.1 Biological roles of R-loops

The past two decades of research revealed that R-loops play important roles in several processes such as gene expression, recombination and genome instability (Santos-Pereira and Aguilera, 2015). Insights into the role of R-loops in gene expression have been gained from genome-wide studies of R-loop occupancy (Chédin, 2016). This has been achieved by DNA-RNA immunoprecipitation followed by high-throughput sequencing (DRIP-seq). The basis of this technique is to pull down the RNA-DNA hybrids using a specific antibody and perform high-throughput sequencing of the immunoprecipitated nucleic acids.

1.2.1.1 R-loops and transcription

R-loops are widely distributed throughout the genome, with a tendency to localize at promoter and terminator regions of RNA Pol II-transcribed genes. They impact on RNA Pol II transcription dynamics at stages like promoter-proximal pausing and favor termination by slowing down RNA Pol II and facilitating the timely recruitment of RNA processing factors (Chan et al., 2014; Chen et al., 2015; Ginno et al., 2012; El Hage et al., 2014; Wahba et al., 2016).

1.2.1.1.1 R-loops and transcription activation

About 60% of human gene promoters contain sequences that are enriched with cytosine-guanine dinucleotides (known as CpG islands) (Jones, 2012). Cytosine methylation at promoter CpG islands is an epigenetic mechanism for transcription inhibition (Deaton and Bird, 2011). For instance, DNA methyltransferase 3B1 (DNMT3B1) silences gene expression by methylating promoter CpGs (Linhart et al., 2007). Interestingly, methylation-resistant CpG islands are characterized by strand asymmetry of guanines and cytosines (a feature termed GC skew). A positive GC-skew (a G-rich template DNA) is strongly correlated with increased R-loop formation at promoter proximal regions (Ginno et al., 2013). Notably, such R-loops block DNMT3B1-dependent silencing of gene expression by preventing promoter CpG methylation (**Fig.3a**) (Ginno et al., 2012). These findings illustrate a role of R-loops that favors transcription initiation.

Another example of the role of R-loops in transcription initiation comes from the regulation of the human *VIM* locus (**Fig.3b**). Anti-sense transcription at the *VIM* locus in colon adenocarcinoma cells forms R-loops at the *VIM* promoter. These R-loops support the opening of the DNA duplex and decrease nucleosome occupancy, thus favoring the binding of transcription factors and activating transcription of the *VIM* gene (Boque-Sastre et al., 2015). A similar mechanism is observed in the budding yeast (*S. cerevisiae*) *GAL* gene cluster. The *GAL* gene cluster is a model to study inducible gene regulation in *S. cerevisiae*, where glucose (repressive) and galactose (active) have counter-acting roles in regulating *GAL* transcription. Notably, R-loops formed between the *GAL* long non-coding RNA and the *GAL* gene enhance *GAL* transcription in the presence of galactose (Cloutier et al., 2016). This is an interesting case suggesting a potential role of environmental cues (sugar in this case) as a trigger of R-loop mediated gene regulation.

1.2.1.1.2 R-loops and transcription termination

It is not yet clear how RNA Pol II transcription termination occurs, mainly because of the poor structural data available so far as a result of the transient nature of termination complexes (Hantsche and Cramer, 2016). However, two models have been widely accepted for RNA Pol II transcription termination. a) The “torpedo model” states that the 5’-3’ exoribonuclease 2 (Xrn2/XRN2, Rat1 in yeast) digests the RNA fragments produced by RNA Pol II downstream of poly-adenylation sites (PAS) displacing RNA Pol II from the template strand through a torpedo-like event (Proudfoot, 1989, 2016). b) The “allosteric model” states that a conformational change in RNA Pol II elongating complexes post-PAS leads to termination (Epshtein et al., 2007; Rosonina et al., 2006). Additionally, the co-existence of both pathways was also suggested (Proudfoot, 2016).

Genome-wide analyses revealed that some mammalian genes are enriched in G-rich pause elements immediately downstream of the PAS. These G-rich elements allow for the pausing of RNA Pol II giving time for Xrn2/XRN2 to cleave the nascent transcript (West et al., 2004). Notably, R-loops are preferentially enriched in human genes at 3’ ends with G-rich pause sites (Skourti-Stathaki and Proudfoot, 2014). Such R-loops regulate transcription termination by pausing RNA Pol II thus giving sufficient time for the recruitment of termination factors (Porrua et al., 2016). For example, the RNA/DNA

helicase − Senataxin (Sen1/SETX) is known to resolve R-loops formed at the gene 3' end. By doing so, Sen1/SETX creates an entry point for Xrn2/XRN2 to access the RNA, facilitating RNA Pol II release as suggested by the torpedo model (Skourti-Stathaki et al., 2011). Consistently in yeast, termination of RNA Pol II generated non-coding RNAs (ncRNAs) requires a termination complex (termination complex NRD; consisting of proteins Nrd1, Nab3 and Sen1 proteins) and cleavage factor (Pcf11 component of cleavage factor 1A) (**Fig.3c**) (Grzechnik et al., 2015). Interaction of the cleavage factor with the termination complex is mediated by the recruitment of Sen1. Ablation of Sen1 disrupts the termination-cleavage complexes and triggers R-loop formation (Grzechnik et al., 2015).

Perturbations in such mechanisms might lead to termination defects letting RNA Pol II escape past transcription-termination region, a process termed transcription read-through. Notably, R-loops can also result from transcription read-through as observed in the *Ube3a* locus, where an R-loop formed in the anti-sense direction affects *Ube3a* gene expression. In this case, transcription read-through of *Snord116* locus (located downstream of *Ube3a*) produces an anti-sense *Ube3a* RNA that inhibits sense *Ube3a* transcription forming a local R-loop (**Fig.3d**) (Huang et al., 2011; Powell et al., 2013).

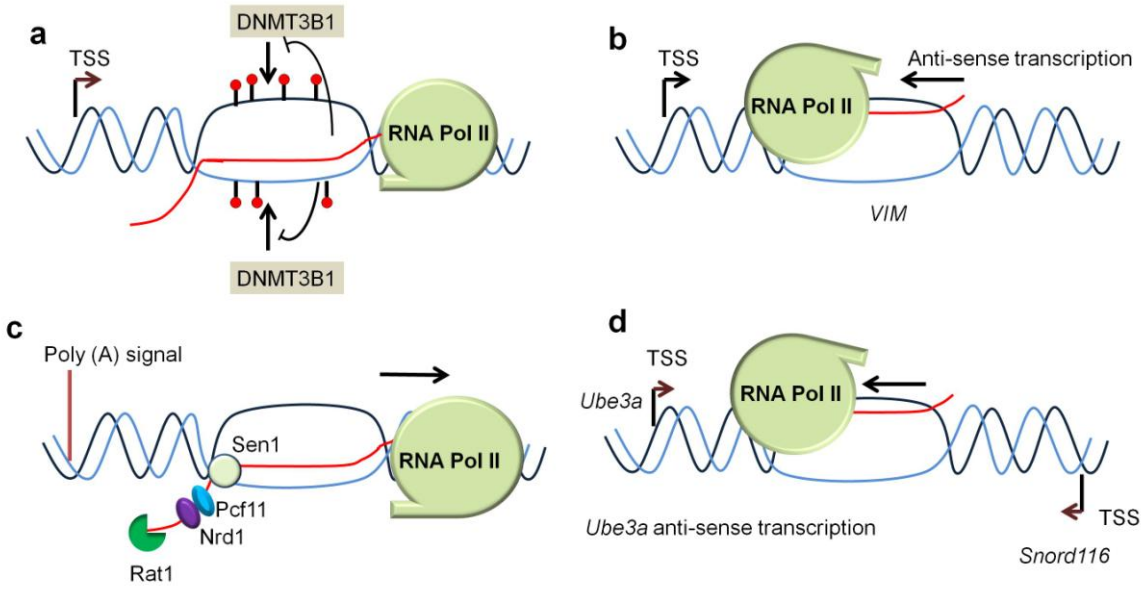


Figure 3. R-loops affect transcription dynamics a) R-loops formed in human CpG island promoters repress methylation (red circles) of genes by preventing binding of DNA methlytransferase B1 (DNMT3B1). **b)** R-loops formed by anti-sense transcription of the *VIM* locus activates transcription in the sense direction by decreasing nucleosome occupancy and promoting transcription factor binding. **c)** Sen1 helicase and Rat1 nuclease work together with other termination factors (Nrd1, Pcf11) to terminate transcription in yeast. **d)** On the contrary, read-through of *Snord116* locus in mouse cells forms R-loops in the anti-sense direction of the *Ube3a* locus thereby inhibiting sense transcription of *Ube3a*.

1.2.1.2 Role of R-loops in chromatin dynamics and genome organization

In addition to a role in gene expression, R-loops can also change the local chromatin structure and regulate the dynamic organization of the genome (Bernstein and Allis, 2005). For example, efficient pausing of RNA Pol II at the human β-actin (*ACTB*) locus is associated with the establishment of dimethylation of lysine 9 of histone H3 (H3K9me2) a repressive chromatin mark (Skourti-stathaki et al., 2014). The human *ACTB* has a G-rich pause site at its 3' end where R-loops form both in sense and anti-sense direction (Skourti-stathaki et al., 2014). Transcription in opposite direction makes two different RNAs that are complementary to each other resulting in the formation of local double-stranded RNA (dsRNA). Eventually, the dsRNA recruits the RNA interference (RNAi) machinery and histone-lysine N-methyltransferase 2 that seeds H3K9me2. All these events force RNA Pol II pausing allowing an efficient transcription termination (Skourti-stathaki et al., 2014). It turns out that H3K9me2 is also indispensable for *C.elegans* development by regulating R-loop levels (Zeller et al., 2016). Similarly, in the fission yeast *S. pombe*, binding of ncRNAs to centromeric chromatin favors local heterochromatization a phenomenon mediated by R-loops with additional requirement of the RNAi machinery (Nakama et al., 2012). Moreover, a prominent increase in condensed chromatin was reported in cells depleted of FAcilitates Chromatin Transcription (FACT) complex and Sen1/SETX in an R-loop dependent manner (Castellano-Pozo et al., 2013).

Eukaryotic chromosomes (which are linear) have a problem in replicating their ends unlike bacterial chromosomes (that are circular). This "end-replication problem", comes from the fact that chromosomes might lose information at their ends after every replication cycle (de Lange, 2009). This is where telomeric sequences come into action; these are repetitive sequences at the end of chromosomes that act as 'buffers' and prevent the loss of genetic material after each round of replication. Telomeres shorten as a function of time and therefore serve as indicators of cellular senescence. Telomerase is the enzyme that replenishes the recurrently shortening telomeres and prevents premature cellular senescence. Eukaryotic telomeres transcribe into non-coding telomeric-repeat-containing RNAs that hybridize with complementary telomeric DNA forming telomeric R-loops in yeast and human cells. R-loops at telomeric regions counteract the telomere shortening in both yeast and human cells that employ telomerase-independent pathways (Arora et al., 2014; Balk et al., 2013; Pfeiffer et al., 2013a).

1.2.1.3 R-loops as drivers of genome instability

On contrary to their function in regulating gene expression and chromatin dynamics unresolved R-loops pose great threats to the genome stability (**Fig.4**). Reasons for this are several: (a) RNA-DNA hybrids are thermodynamically more stable than double-stranded (dsDNA) (Thomas et al., 1976) adopting a conformation between the B-form of dsDNA and the A-form of dsRNA (Lesnik and Freier, 1995), making it hard for cells to resolve these hybrids. (b) Additionally, a RNA-DNA hybrid in an R-loop, displaces a ssDNA exposing it to DNA damage agents, such as nucleases and enzymes that cause DNA single-strand breaks (SSBs) (Paulsen et al., 2014; Sollier and Cimprich, 2015), which can eventually develop into DNA DSBs (Hatchi et al., 2015; Khoronenkova and Dianov, 2015). (c) The R-loop potentially prevents nucleosome assembly and creates a damage-prone 'naked' DNA region (Dunn and Griffith, 1980; Sollier and Cimprich, 2015). (d) R-loops stall the transcription machinery, which can in turn block replication complexes causing transcription-replication collisions. In fact, hindrance of replication is thought to represent the most common cause of R-loop dependent DNA damage (Helmrich et al., 2013). For instance, R-loops at immunoglobulin class-switch regions cause replication fork stalling and chromosomal rearrangements (Gaillard et al., 2015; Helmrich et al., 2013).

The role of R-loops in promoting DNA damage was first demonstrated in yeast and later in human cells lacking the TRanscription EXport complex (THO/TREX*)* (Domínguez-Sánchez et al., 2011; Huertas and Aguilera, 2003). THO/TREX is a multi-protein complex involved in transcription and RNA export. Depletion of THO/TREX components leads to aberrant R-loop formation and DNA damage (Domínguez-Sánchez et al., 2011). Interestingly, *C. elegans thoc-2* mutants exhibit replication-fork stalling and occurrence of DNA breaks, suggesting that *thoc-2* mutants go through severe replication stress (Pfeiffer et al., 2013b, Castellano-Pozo et al., 2012, Santos-Pereira et al., 2014). Similarly depletion of RNA processing factors like serine/arginine splicing factor 1 (SRSF1, also known as ASF/SF2), Aquarius (AQR) and SETX challenges the genome integrity via R-loops (Li and Manley, 2005; Paulsen et al., 2014; Yüce-Petronczki and West, 2012).

Nevertheless, DNA damage itself can also lead to the formation of R-loops. For instance, DNA lesions within transcriptional units cause RNA Pol II pausing and local displacement of RNA processing factors, which create opportunities for DNA invasion by the nascent RNA and the resulting R-loop formation. Strikingly R-loops can also trigger the DNA damage response (DDR) creating a feedback loop between R-loops and the DDR (Tresini et al., 2015). Moreover, camptothecin, a topoisomerase I inhibitor that leads to RNA Pol II pausing and promotes R-loops also causes DNA damage (Baranello et al., 2009; Khobta et al., 2006; Liu et al., 1996; Marinello et al., 2013, 2016; Paulsen et al., 2014). This further illustrates the coupling between DNA damage, R-loops and transcription dynamics.

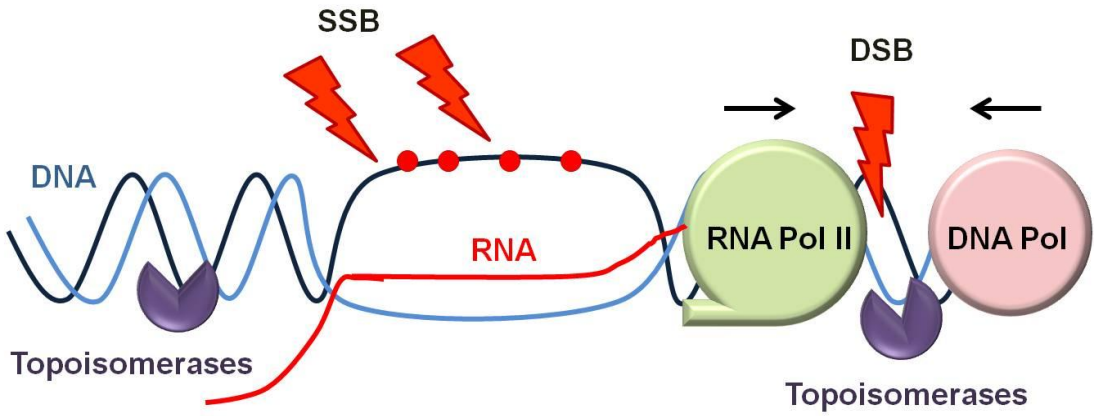


Figure 4. R-loops are a major source of genome instability. R-loops pose great threat to genome integrity by interfering with RNA Pol II and DNA Pol creating transcription-replication collisions leading to double strand breaks (DSBs). Additionally, exposed ssDNA in a R-loop can act as a bait for DNA damaging enzymes and mutagenic agents creating single strand breaks (SSBs) that can eventualy frutify as DSBs. Further activity from Topoisomerases that try to relieve negative and positive supercoiling during replication and transcription can also act as source of genome instability.

1.3 R-loops and disease

Imbalance in R-loop homeostasis has been associated primarily with human neurological disorders and as a possible cause of genome instability in cancer cells (**Fig.5**).

1.3.1 Role of R-loops in neurological disorders

R-loops are implicated in the pathogenesis of neurodegenerative diseases like Amyotrophic lateral sclerosis (ALS), Aicardi-Goutières Syndrome (AGS) or Fragile X syndrome. Recent high-throughput R-loop sequencing in fibroblasts from AGS patients shows that R-loops are a source of unprocessed nucleic-acids burden seen in these cells (Lim et al., 2015). ALS is a severe debilitating neurodegenerative disease linked to mutations in various genes implicated in RNA metabolism (Hill et al., 2016; Salvi and Mekhail, 2015). Genetic and mutational studies helped to reveal connections between various ALS-linked factors and R-loop regulation (Groh and Gromak, 2014). Another neurological disease associated with R-loop biology is Fragile X syndrome and Fragile X-associated tremor/ataxia syndrome (Colak et al., 2014; Groh et al., 2014). The reduced expression of the Fragile X mental retardation 1 (*FMR1*) gene containing CGG repeats in the 5' UTR is the primary cause of the syndrome. Notably, *FMR1* gene silencing is mediated by co-transcriptional R-loops formed at expanded CGG repeats (Groh et al., 2014).

R-loops tend to form in repetitive DNA sequences (like in ribosomal DNA and centromeric regions), because of the abundant complementarity between the nascent RNA and the template DNA (Nadel et al., 2015). For example, transcription of CTG repeat sequences leads to R-loop formation that stimulates repeat instability in bacteria and human cells (Lin et al., 2010; Reddy et al., 2011). Tri-nucleotide expansions might also impair transcription, as illustrated in Friedreich's ataxia, a common inherited ataxia. In Friedreich's ataxia unstable GAA repeat expansions in the first intron of the Frataxin gene reduce its expression by a mechanism involving the formation of R-loops and RNA Pol II stalling at the repeat sequences (Grabczyk et al., 2007; Reddy et al., 2011). Similarly, R-loops formed in hexanucleotide repeat expansions-GGGGCC in the chromosome 9 open reading frame 72 are implicated in ALS (Haeusler et al., 2016).

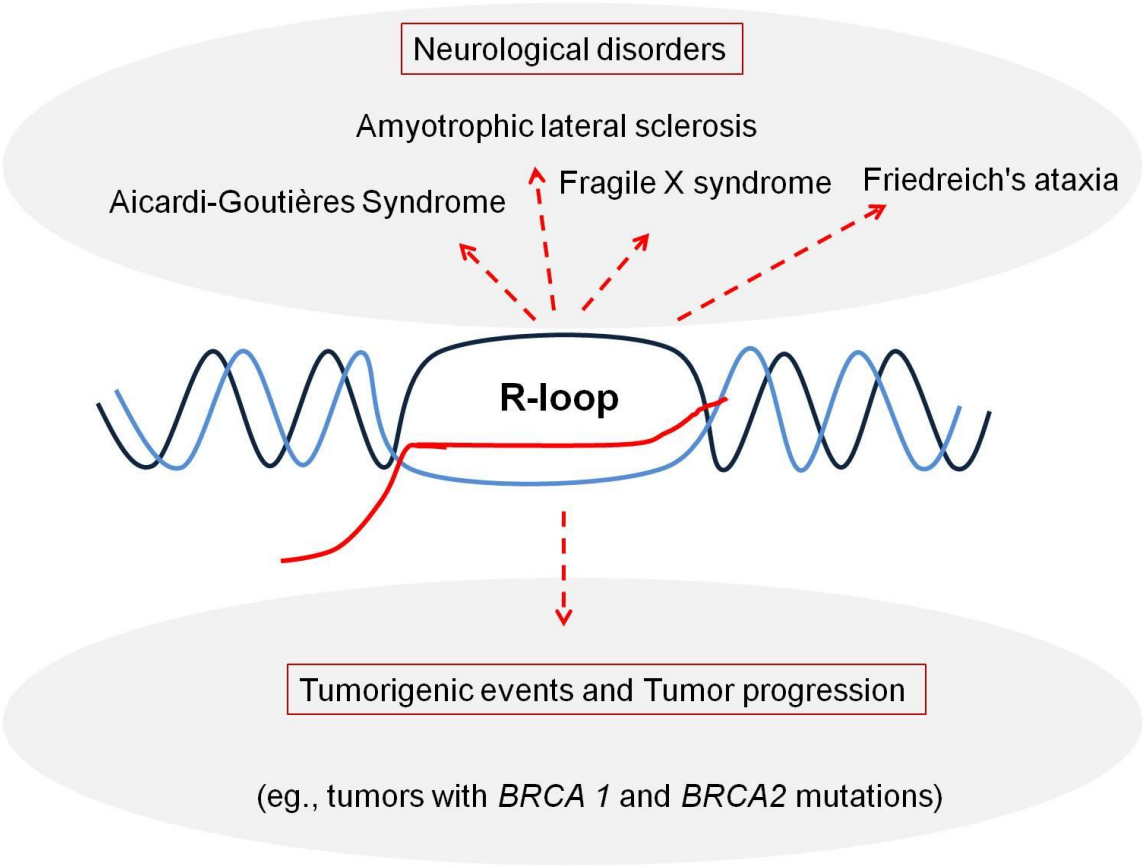


Figure 5. Role of R-loops in human pathologies. R-loops play a crucial role in several neurological disorders (e.g. AGS, ALS and Fragile X syndrome). Mutations in genes coding for proteins linked to R-loop homeostasis are frequently involved in the development of these neurological disorders. R-loops can also favor tumorigenic events that drive cancer progression and can contribute to the high levels of genome instability observed in most cancers.

1.3.2 Role of R-loops in cancer

Evidence is gaining momentum to support the role of R-loop-mediated genome instability as a driver for cancer progression. For example, tumor suppressor genes like *BRCA1* and *BRCA2* prevent R-loop accumulation (Bhatia et al., 2014; Hill et al., 2014). Both *BRCA1*$^{-/-}$ and *BRCA2*$^{-/-}$ cells accumulate R-loops and DNA DSBs. However, how exactly BRCA1 and BRCA2 regulate R-loop levels it still unclear. One mechanism BRCA1 employs to regulate R-loop levels is by orchestrating the mRNA splicing machinery and changing splicing patterns (Savage et al., 2014). On the other hand, BRCA2 is required to protect stalled replication forks and prevent their collapse, suggesting that BRCA2 might prevent R-loop accumulation by allowing replication fork to restart after cells clear R-loops (Berti and Vindigni, 2016). Malfunction of one or both of these proteins might increase the R-loop burden in cancer cells triggering mutations favoring cancer progression.

An interesting example of R-loop driven tumorigenic process is observed in Burkitt lymphoma, a cancer characterized by a translocation between the proto-oncogene *Myc* and immunoglobulin switch regions. Notably, the translocation sites are predominantly GC rich, a feature known to favor R-loop formation (Ramiro et al., 2004; Ruiz et al., 2011; Pefanis and Basu, 2015). Another example of the impact of R-loops in cancer development is the sequestration of the THO/TREX complex by the viral protein ORF75. Evidence suggests that this is a tumorigenic mechanism in cells infected with Kaposi's-sarcoma-associated herpes virus (Jackson et al., 2014). In fact, these cells accumulate increased R-loops and genome instability.

1.4 Cellular mechanisms to maintain R-loop homeostasis

Cells accumulate R-loops naturally, but they also evolved various mechanisms to keep R-loop levels constantly in check to prevent genome instability. Cellular mechanisms that maintain R-loop homeostasis fall into two major classes: a) those that prevent the accumulation of R-loops and b) those that resolve existent R-loops (**Fig.6**).

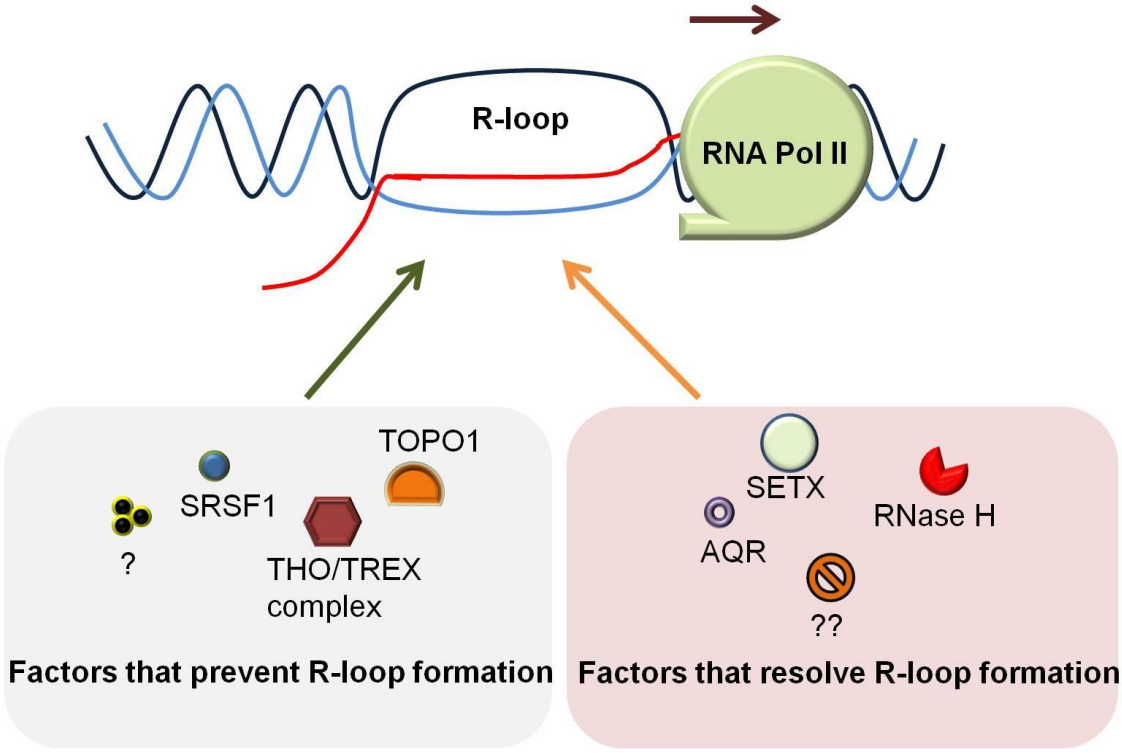


Figure 6. Molecular mechanisms to regulate R-loop homeostasis. Cells can regulate R-loop homeostasis by employing mechanisms to prevent *de novo* R-loop formation or resolving existent ones. RNA processing factors like serine-arginine splicing factor 1 (SRSF1), RNA transcription and export complex (THO/TREX) and other RNA processing machinery prevent R-loop formation by sequestering nascent RNA from binding to template DNA strand. Additionally, topoisomerases (eg., TOPO1) can also prevent R-loop formation by relieving the negative supercoiled DNA. Cells also employ ribonucleases H (RNase H) and DNA/RNA helicases like aquaris (AQR), senataxin (SETX) and other factors to either digest or resolve R-loops.

1.4.1 Mechanisms to prevent deleterious R-loops

Cells can prevent R-loop formation by two possible mechanisms. 1) Stabilizing the native dsDNA conformation during replication and transcription and 2) coating the nascent RNA with proteins that physically obstruct the hybridization of RNA with DNA.

Unwinding of the dsDNA during transcription generates positive DNA supercoils ahead and negative DNA supercoils behind the transcription complexes (Liu and Wang, 1987). This torsional strain is relieved by toposimerases during transcription (and replication) (Champoux, 2001; Lee and Young, 2000). As negatively supercoiled dsDNA is a critical factor for R-loop formation (Chédin, 2016) topoisomerases prevent co-transcriptional R-loops by relieving negative supercoiled DNA behind elongating complexes (Yang et al., 2014; Wilson-Sali and Hsieh, 2002). For instance, yeast topoisomerase 1 and 2 can regulate R-loops in ribosomal DNA, depletion of which increases R-loop levels and causes RNA polymerase I (RNA Pol I) stalling (El Hage et al., 2010). Similarly, topoisomerase I (TOPI) suppresses genomic instability in mammalian cells by preventing co-transcriptional R-loops and reduces conflicts between the transcription and replication machineries (Tuduri et al., 2009).

A growing body of evidence suggests that RNA processing factors make a substantial contribution in maintaining R-loop homeostasis by preventing hybridization between the nascent RNA and the template DNA (Montecucco and Biamonti, 2013). Most of the RNA processing (i.e., capping, splicing, polyadenylation and folding) takes place co-transcriptionally (Bentley, 2014). RNA processing proteins bind to nascent RNA forming a ribonucleoprotein (RNP) complex. In such a secluded environment, the nascent RNA is physically forbidden to invade the template DNA and form an R-loop. Therefore, deficiency of RNA processing factors and/or RNP biogenesis leads to R-loop formation. For instance, depletion of the already mentioned THO/TREX causes the accumulation of R-loops (Domínguez-Sánchez et al., 2011; Huertas and Aguilera, 2003; Pfeiffer et al., 2013a). Actually, this illustrates the first link established between RNP biogenesis and R-loop formation (Santos-Pereira and Aguilera, 2015). Depletion of the splicing factor SRSF1 in both chicken DT-40 and HeLa cells also induces R-loop-mediated genome instability possibly by perturbing RNP formation (Li and Manley, 2005; Paulsen et al., 2014). Thus, the dynamics of co-transcriptional RNP formation is a critical determinant of R-loop homeostasis. Importantly, recruitment of the RNA-binding proteins, such as the

spliceosome components, to the nascent RNA, but not the process of splicing itself is critical to prevent R-loop formation and genome instability (Bonnet et al., 2017).

1.4.2 Mechanisms to resolve deleterious R-loops

To restore the native conformation of the duplex DNA, cells have to either digest or unwind the RNA moiety of an R-loop. For this purpose cells use nucleases or helicases, enzymes that either digest or unwind RNA from the R-loop, respectively (Sollier and Cimprich, 2015). Perhaps the most widely characterized nuclease that digests RNA within an R-loop is ribonuclease H (RNase H). Unlike other ribonucleases, RNase H specifically digests RNA that is hybridized to DNA (Schultz and Champoux, 2008). Two types of RNase H enzymes are present in eukaryotes: RNase H1 and RNase H2. Both types of RNase H enzymes can digest RNA from R-loops, although they have different physiological roles (Schultz and Champoux, 2008). RNase H1 is located in the nucleus and in mitochondria of mammalian cells, and digests R-loops that arise from nuclear and mitochondrial transcription (Cerritelli et al., 2003). RNase H2 digests RNA in RNA-DNA hybrids formed during replication (Sparks et al., 2012). RNase H2 can also remove single ribonucleotide moieties mis-incorporated in DNA, whereas RNase H1 works on longer patches of RNA-DNA hybrids (Zimmer and Koshland, 2016). Nevertheless, both enzymes are required for cellular survival. In agreement, *Rnaseh1* null mice exhibit embryonic lethality (Cerritelli et al., 2003). Moreover, both depletion or constitutive activation of these enzymes is deleterious to cell survival, revealing the importance of a tight regulation of their cellular levels (Cerritelli and Crouch, 2009). For instance, deletion of RNase H1 stabilizes RNA-DNA hybrids around DNA DSB sites impairing recruitment of the homologous recombination machinery (Ohle et al., 2016). On the other side, overexpression of RNase H1 destabilizes hybrids and leads to severe loss of repetitive regions around DSBs (Ohle et al., 2016). Furthermore, mutations in RNase H2 are strongly correlated with high levels of R-loops in fibroblasts from AGS patients (Lim et al., 2015).

Besides RNase H, cells can employ RNA/DNA helicases to unwind RNA from R-loops allowing DNA to re-anneal into a dsDNA (Costantino and Koshland, 2015). RNA/DNA helicases like SETX, bacterial RecG DNA helicase, Rho transcription termination factor, yeast DNA helicase Pif1, human RNA helicase DEAH box protein 9 and human RNA

helicase aquarius (AQR) have all been shown to maintain R-loop homeostasis (Harinarayanan and Gowrishankar, 2003; Hong et al., 1995; Paulsen et al., 2014). As already mentioned, depletion of the RNA/DNA helicase SETX leads to increased occurrence of R-loops in transcription termination sites (Becherel et al., 2013; Kim et al., 1999; Mischo et al., 2011; Skourti-Stathaki et al., 2011). The DEAH box protein 9 was shown to specifically resolve R-loops structures *in vitro* (Chakraborty and Grosse, 2011). In addition, yeast Pif1 helicase might reduce R-loop levels by destabilizing G4 DNA (G quadruplexes or G quadrets) structures allowing duplex DNA formation (Boué and Zakian, 2007; Mendoza et al., 2015). G4 DNA is a tertiary nucleic acid structure formed in guanine rich regions. A minimum of 4 guanine bases can form a square planar structure through hoogsten pairing called a guanine tetrad. Two or more guanine tetrads stack on top of each other to form a G4 DNA (Maizels and Gray, 2013). Interestingly, R-loops tend to form with transcription of C-rich template DNA forming a G-rich RNA and displacing a G-rich ssDNA (Ginno et al., 2013). Such, G-rich ssDNA is likely to form a G4 structures (Sollier and Cimprich, 2015). Therefore, helicases like yeast Pif1 that relieve G4-DNA on ssDNA can indirectly regulate R-loop levels. However, despite their established role in resolving R-loops, how the function of these enzymes is regulated needs to be clarified individually.

1.5 SRPK1 and SRPK2 as central modulators of RNA processing factors

Protein phosphorylation is a master regulatory mechanism that operates in virtually all cellular processes, including RNA processing or the DDR. For instance, during the DDR, phosphorylation impacts on the localization, stability and activity of several proteins, acting as signal cues to sense, transduce and repair the DNA lesion. In fact, 900 phosphorylation events involving about 700 proteins are estimated at the onset of the DDR (Matsuoka et al., 2007). These numbers emphasize the important role of protein kinases (Lahiry et al., 2010) with a special function in maintaining genome stability (Jackson and Bartek, 2009). In addition, recent proteomic study revealed that serine/arginine protein kinases (SRPKs; that include SRPK1 and SRPK2) are required to orchestrate distinct molecular events involved in RNA processing (Varjosalo et al., 2013). Indeed, a major fraction of SRPK1 and SRPK2 substrates have roles related with transcription and RNA processing (**Fig.7a**). These substrates range from splicing factors to several helicases involved in maintaining RNA homeostasis (Varjosalo et al., 2013).

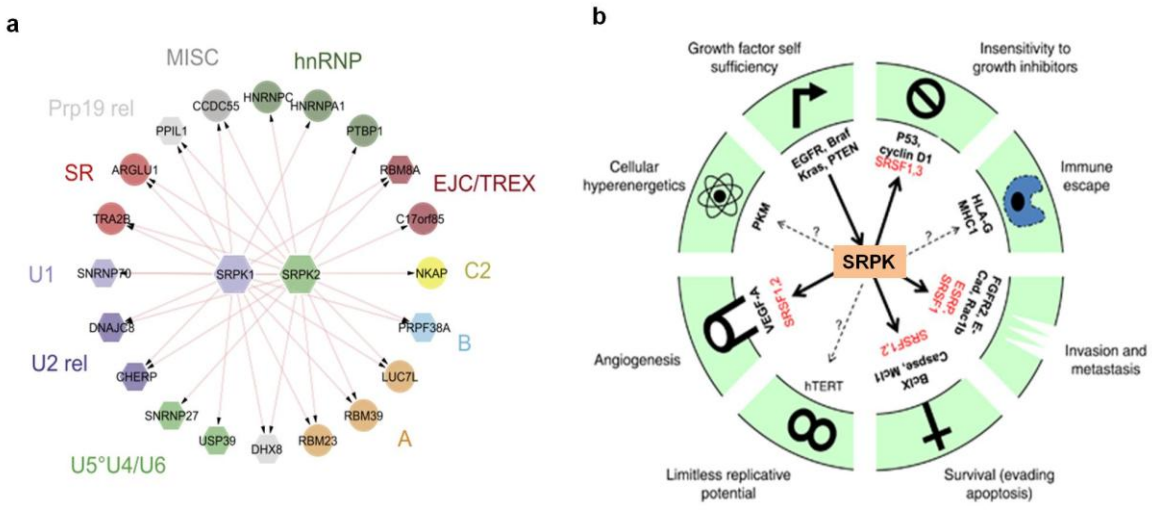


Figure 7. SRPKs at the centre of cancer-related mechanisms and RNA processing. a) SRPK1 and SRPK2 phosphorylate various SR proteins that play critical role in different steps of RNA processing such as splicing and export. Adapted from Varjosalo et al., 2013. **b)** SRPKs lie at the centre of alternative aplicing decisions that are detrimental for various cellular processes like proliferation, growth and apoptosis. Dysregulation of SRPKs changes splicing patterns promoting cancer development. Adapted from Oltean and Bates, 2013.

1.5.1 Functions of SRPK1 and SRPK2

SRPK1 and SRPK2 share a sequence homology of ~57%, but have distinct roles by modulating splicing decisions of crucial biological functions (**Fig.7b**) (Wang et al., 1998). For instance, chemical inhibition of SRPK1 or *SRPK1* RNAi was found to effectively block angiogenesis by changing the alternative splicing pattern of vascular endothelial growth factor (a pro-angiogenic molecule) (Amin et al., 2011). Additionally, the *C. elegans* orthologue of SRPK1 and human SRPK1 are essential for embryogenesis of *C. elegans* and spermatogenesis, respectively (Galvin et al., 2011; Papoutsopoulou et al., 1999). However, SRPK2 plays an important role in regulating alternative splicing in response to DNA damage in neuronal cells (Vivarelli et al., 2013; Wang et al., 1998).

While SRPK1 is part of the U1 small nuclear ribonucleoprotein (U1 snRNP) complex of the splicing machinery, SRPK2 is part of the U4/U6.U5 tri-snRNP spliceosomal complex (Kamachi et al., 2002; Mathew et al., 2008). Both play crucial roles in constitutive and alternative splicing. As the name suggests, SRPK1 and SRPK2 phosphorylate serine residues of serine/arginine rich proteins (SR proteins), a group of proteins bearing a arginine/serine (RS) domain either in their N or C terminus (Zhou and Fu, 2013). SRPKs can phosphorylate up to 8 serines in the RS domain using a processive mechanism in which the kinase stays attached to the substrate after each round of phosphorylation. The remaining serines in the RS domain are modified in a distributive manner in which the kinase and substrate dissociate after each phosphorylation event. SRPKs tend to have a preference for serines adjacent to arginines, rather than threonines (Ghosh and Adams, 2011; Stojdl and Bell, 1999).

SR proteins have diverse roles in cellular process like cell signaling, cell cycle progression and notably RNA and DNA metabolism (Giannakouros et al., 2011). SR proteins that play a role in RNA processing have a RS domain and one or more RNA binding motifs (Sapra et al., 2009). The RS domain drive protein-protein interactions, which in turn facilitate the appropriate binding of the RNA binding motifs to *cis*-regulatory elements of pre-messenger RNA (pre-mRNA) (Manley and Tacke, 1996). The RS domain of spliceosomal SR proteins is profusely phosphorylated and the splicing activity is significantly dependent on the degree and timing of the phosphorylation event (Cazalla et al., 2002). *In vitro* and *in vivo* studies strongly suggest that the phosphorylation/de-phosphorylation switch leads to changes in the protein interaction milieu of the spliceosome and affects splicing (Long and

Caceres, 2009; Stojdl and Bell, 1999; Xiang et al., 2013). Thus, SRPK-mediated phosphorylation of SR proteins plays a crucial role in processing nascent RNA.

1.5.2 Regulation of SRPK1 and SRPK2

SRPK1 and SRPK2 are conserved kinases found in yeast, worms, slime molds, plants and mammals. Structurally they have a typical bilobal domain found in all eukaryotic protein kinases separated by an insert domain (Ghosh and Adams, 2011). The insert domain is important in regulating the sub-cellular localization and in turn the function of SRPK1 and SRPK2. The insert domain acts as a docking site for the binding of chaperone proteins (Zhong et al., 2009). In mammalian cells, different environmental cues and stress signals affect SRPKs sub-cellular localization by regulating the binding of chaperone proteins to the insert domain (Ghosh and Adams, 2011).

Additionally, the function of SRPK1 and SRPK2 is regulated by post-translational modifications. For instance, protein kinase B (AKT) mediated threonine-492 phosphorylation of SRPK2 promotes its nuclear translocation leading to cyclin D1 up-regulation and cell cycle re-entry of neuronal cells leading to their apoptosis (Jang et al., 2009). AKT is the only known upstream kinase of SRPKs and creates diversity in SRPK1 and SRPK2 functionality. While AKT-SRPK2 axis works in regulating neuronal apoptosis, AKT-SRPK1 axis regulates alternative splicing decisions in angiogenesis of gliomas and in malignancy of hepatocellular carcinoma (Wang et al., 2014; Wu et al., 2014; Zhou et al., 2013). The 14-3-3 protein counteracts AKT phosphorylation of SRPK2 by sequestering SRPK2 (Jang et al., 2009).

1.5.3 Role of SRPK1 and SRPK2 in human malignancies

Mutations and differential expression of SRPKs have been observed in various diseases like cancer, neurological pathologies and viral infections (Giannakouros et al., 2011).

1.5.3.1 SRPKs in cancer

A recent phospho-proteomic screen in head and neck squamous cell carcinoma cells identified a hyperphosphorylated form of SRPK2 regulating splicing patterns (Radhakrishnan et al., 2016). Inhibition of SRPK2 significantly decreased the colony forming capability of head and neck squamous carcinoma cells. Additionally, SRPK2 and SR-like protein acinus are overexpressed in some human acute myelogenous leukemia patients. Phosphorylation of acinus by SPRK2 elevates cyclin A1 expression levels and increase leukemia cell proliferation (Hong et al., 2011; Jang et al., 2008). In non-small cell lung cancer, the ectopic expression of SRPK1 promotes the growth and migration of cancer cells, whereas *SRPK1* knock-down inhibits tumor growth, migration, and tumorigenicity (Liu et al., 2016). Upregulation of SRPK1 was also correlated with high metastasis rates and poor survival in breast cancer patients (Lin et al., 2014; Wang et al., 2014).

Additionally, SRPKs have been implicated in viral pathologies. For instance, the SRPK inhibitor SRPIN340 suppresses hepatitis C virus replication *in vitro* in a dose-dependent manner (Karakama et al., 2010). Additionally, phosphorylation of the SR splicing factor 4 by SRPK2 increased to 20-fold under infection. Chemical inhibition of SRPK1 and SRPK2 using SRPIN340 inhibited viral replication (Fukuhara et al., 2006) suggesting that SRPKs lie at the heart of the host splicing machinery hijacked by viruses. However, it is still not clear if SRPKs play a role in viral induced tumorigenic process

1.5.3.2 SRPKs in neuronal pathologies

A growing body of evidence has implied SRPKs in neuronal pathologies (Chan and Ye, 2013). For instance, threonine 492-phosphorylated SRPK2 binds 14-3-3 protein and triggers cell cycle progression and cell death in terminally differentiated neurons (Jang et

al., 2009). Additionally, phosphorylation of Tau by SRPK2 was linked to Alzheimer's disease (Hong et al., 2012). Tau proteins are soluble microtubule-associated proteins expressed specially in neuronal cells. Deregulation of Tau are causally linked to Alzheimer's disease and other neuronal degeneration (Ballatore et al., 2007), suggesting a role of SRPK2 in additional neuronal pathologies. SRPK1, but not SRPK2, translocates from the nucleus into the cytoplasm under ischemic stress in neurons (Erdö et al., 2004). This differential behavior may relate with the findings of an anti-apoptotic role of SRPK1 in contrast to the pro-apoptotic role of SRPK2 in neurons (Hong et al., 2011; Kamachi et al., 2002).

1.6 Open questions

Despite the already known role in splicing, the role of SRPKs in regulating the genome stability (a hallmark of cancer) is yet to be elucidated. We therefore asked the following questions:

1) Do SRPKs have any role in maintaining R-loop mediated genome stability?

   Rationale: This question stems from the fact that the proper functioning of RNA processing factors (many of them are SR proteins) is crucial to maintain cellular R-loop levels.

   Hypothesis: We hypothesize that SRPKs regulate R-loop levels by fine tuning the function of several RNA processing factors.

2) How do SRPKs regulate R-loop homeostasis?

   Rationale: Various RNA binding proteins and RNA/DNA helicases are substrates of SRPKs anticipating a putative role of SRPK–dependent phosphorylation in maintaining R-loop homeostasis.

   Hypothesis: We hypothesize that SRPKs regulate R-loop levels by driving the phosphorylation of RNA processing factors. Failure of SRPKs activity might lead to aberrant processing of the nascent RNA transcript and trigger R-loop formation. SRPKs may balance R-loop levels by either preventing the formation of deleterious R-loops or by driving the resolution of the existing ones.

3) Is the link between SRPKs and R-loop mediated genome instability relevant for cancer development?

   Rationale: Both *SRPK1* and *SRPK2* are mutated in several cancers such as breast, colon, pancreas, head/neck and leukemia.

   Hypothesis: We hypothesize that SRPKs are major regulators of cellular R-loop levels and the genetic inactivation of SRPKs drives tumorigenic processes.

Goal

The main goal of this thesis is to provide mechanistic insights into each of the aforementioned questions and to clarify the physiological relevance of the scrutinized processes.

A single twig breaks, but the bundle of twigs is strong.

Tecumseh

RNA, "I may be weak, feeble and fragile by myself, but with my twin, I am hard to break"

Anonymous

Chapter 2.
The role of SRPK2 in RNA-mediated genome instability

Highlights

- SRPK1 and SRPK2 are associated with RNA Pol II complexes.
- Depletion of SRPK2 induces DNA damage.
- R-loops are the major source of DNA damage in *SRPK2*-depleted cells.

2.1 SRPK2 is necessary to protect the genome integrity

SRPK1 and SRPK2 orchestrate the phosphorylation of several SR proteins acting as RNA processing factors. As RNA processing factors play a critical role in maintaining genome stability (Paulsen et al., 2009) we asked whether SRPKs function to prevent DNA damage. To gain novel insights into the biological function of SRPK1 and SRPK2 in DNA damage, we depleted either *SRPK1* or *SRPK2* from U-2OS and HeLa cells by RNAi and measured the amount of resulting DNA damage by immunofluorescence (**Fig.8a**) and western blot (**Fig.8b**) of phosphorylated histone H2AX (γH2AX) – an hallmark of the DNA damage response to DNA DSBs, here used as a proxy for DNA damage. *SRPK1* and *SRPK2* were depleted using different pools of small interfering RNAs (siRNA). As a control we included a pool of siRNA directed against firefly luciferase gene. Following the depletion of *SRPK2*, but not of S*RPK1*, there was a significant increase in total cellular γH2AX levels (**Fig.8b**). Additionally, we also observed a striking increase in nuclear γH2AX foci in replicating cells (revealed by cyclin A staining) depleted of *SRPK2* (**Fig.8a**).

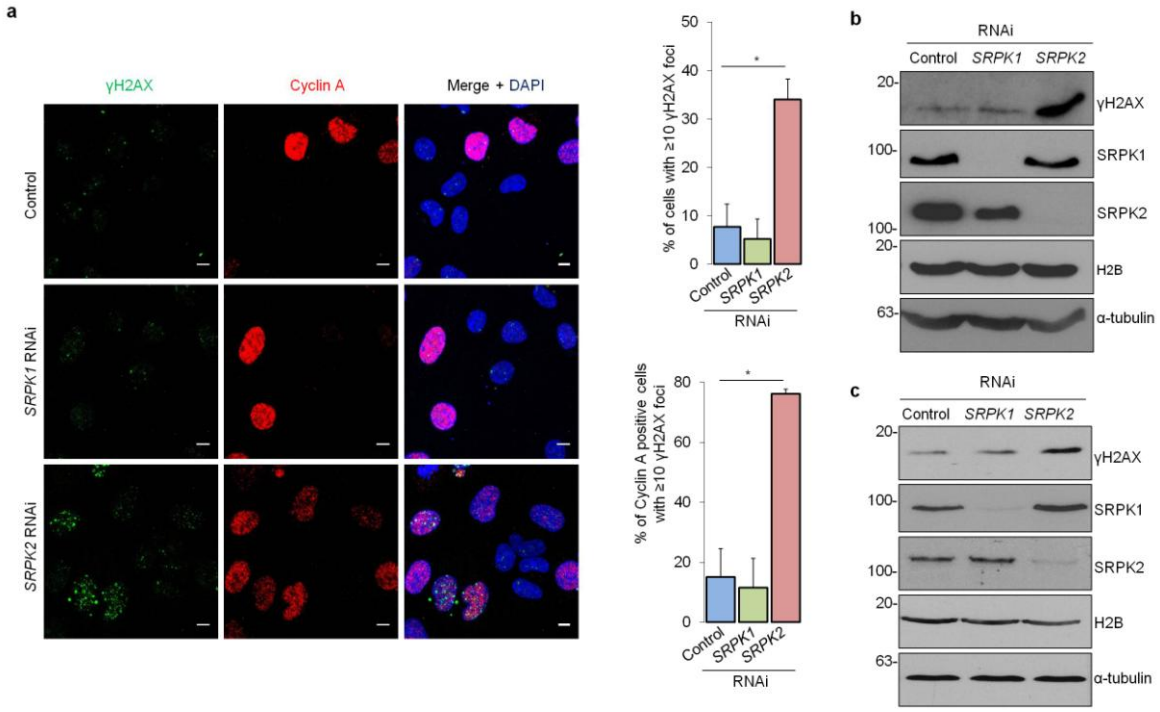


Figure 8. SRPK2 is necessary to maintain the genome integrity. a) γH2AX foci (green) and cyclin A (red) staining in control, *SRPK1* and *SRPK2* RNAi-depleted U-2OS cells. Means and standard deviations of the percentage of cells with 10 or more γH2AX foci were plotted. At least 100 cells from three independent experiments were scored. Scale bars:10 μm. Statistical significance was determined using two-tailed Student's t-test *p<0.05. **b)** Immunoblots showing γH2AX, SRPK1 and SRPK2 levels in U-2OS cells upon depletion of *SRPK1* or *SRPK2* by RNAi. Histone H2B and α-tubulin served as loading controls. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments. **c)** Immunoblots showing γH2AX, SRPK1 and SRPK2 levels in HeLa cells upon depletion of *SRPK1* or *SRPK2* by RNAi. Histone H2B and α-tubulin served as loading controls. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments.

2.2 RNA Pol II acts as a molecular bridge for the association of SRPK1 and SRPK2 to chromatin

Within the cell, SRPK1 and SRPK2 can be detected both in the cytoplasm and in the nucleus where they adopt a speckles-like distribution pattern, which is typical of splicing proteins (**Fig.9a**). Additionally, we also observed both SRPK1 and SRPK2 in the nucleoplasm and associated with chromatin as revealed by sub-cellular fractionation (**Fig.9b**). Intrigued by the chromatin association of SRPKs, we reasoned that either RNA Pol II or the RNA transcript served as a linker between SRPKs and chromatin.

In order to inspect the role of RNA in bridging SRPK1 and SRPK2 to chromatin we isolated chromatin from RNase A-digested nuclei. Digestion of nuclear RNA using RNase A before chromatin isolation released the RNA-bound U1-A30K (splicing factor) from the chromatin. U1-A30K acted as control for RNase A digestion. We observed that the RNase A digestion did not perturb the interaction of SRPK1 and SRPK2 with chromatin, suggesting an RNA-independent interaction (**Fig.9c**).

We then reasoned that SRPKs could be tethered to chromatin via RNA Pol II. We further hypothesize that the C-terminal domain (CTD) of RNA Pol II largest subunit RPB1 is directly implicated in the interaction with SRPKs. The RNA Pol II CTD consists of multiple repeats of the YSPTSPS heptad, varying in number from 26 in yeast to 52 in human. The RNA Pol II CTD is subject to extensive post-translational modification, notably phosphorylation, during the transcription cycle. The RNA Pol II CTD also acts as a molecular platform to recruit multiple RNA processing factors during transcription (de Almeida and Carmo-Fonseca, 2012; Zaborowska et al., 2016; Harlen et al., 2016). To this end, we immunoprecipitated endogenous SRPK1, SRPK2 and RNA Pol II from isolated nuclear extracts (**Fig.9e**). Further, DNase I digestion was performed to exclude DNA that might act as linker between SRPKs and RNA Pol II (**Fig.9d**). This experiment revealed that both SRPK1 and SRPK2 are associated with RNA Pol II complexes, suggesting that RNA Pol II act as a molecular bridge that brings SRPKs to chromatin (**Fig.9e**).

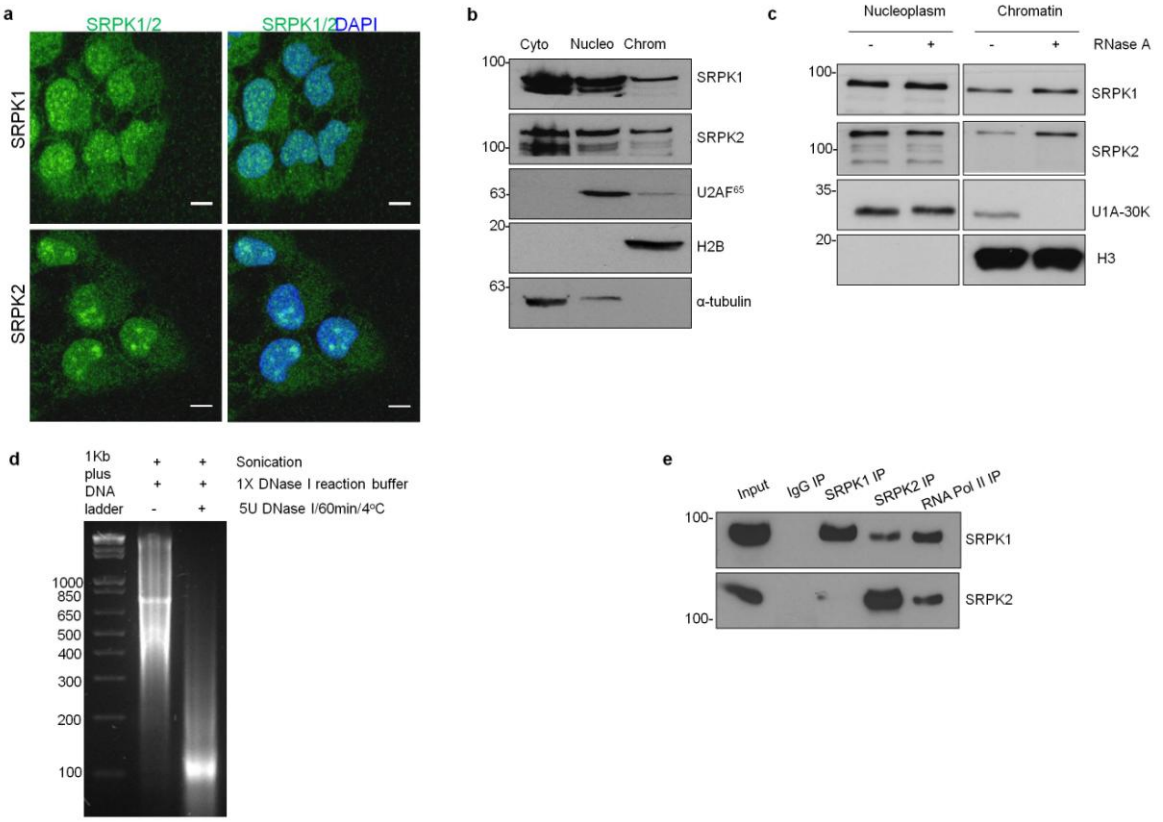


Figure 9. SRPK1 and SRPK2 are chromatin associated proteins. a) Immunofluorescence analysis of endogenous SRPK1 and SRPK2 in U-2OS cells. Scale bars: 10 μm. **b)** Immunoblots showing SRPK1 and SRPK2 levels in cytoplasm (Cyto), nucleoplasm (Nucleo) and chromatin (Chrom) biochemical fractions of U-2OS cells. α-tubulin, $U2AF^{65}$ and histone H2B immunoblots served as controls for independent fractions. Data are representative of three independent experiments performed. **c)** Immunoblots showing the SRPK1 and SRPK2 levels in the nucleoplasm (Nucleo) and chromatin (Chrom) fractions treated with/without RNase A in HeLa cells. U1A-30K and histone H3 immunoblots served as controls for RNase A treatment and fractionation protocol respectively. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments performed. **d)** Genomic DNA from MCF7 cells run on 1% agarose gel with (+) and without (-) DNase I digestion. Note that DNase I digests genomic DNA to ~100bp fragments. 1Kb plus DNA ladder are shown on the left in base pairs. **e)** Immunoprecipitations of SRPK1, SRPK2 and RNA Pol II in MCF7 cells nuclear extracts. Purified complexes were resolved by SDS-PAGE and blotted with antibodies against SRPK1 and SRPK2. The Input lane represents 10% total cell lysates and IgG IP denotes the negative control immunoprecipitation. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments.

Given the association of SRPK1 and SRPK2 with RNA Pol II, we then asked, if inhibition of transcription could interfere with the DNA damage observed in *SRPK2*-depleted cells. To this end, we inhibited RNA Pol II transcription initiation using triptolide (Titov et al., 2011). Indeed, inhibition of transcription reduced the number of nuclear γH2AX foci in *SRPK2*-deficient cells (**Fig.10a-c**). The 70 minutes incubation time with triptolide in our experimental setting did not produce any noticeable cell cycle alterations (**Fig.10d**) suggesting that the effect on DNA damage is driven by the transcription inhibition. This data suggest that active RNA Pol II transcription is necessary to induce DNA damage in *SRPK2*-depleted cells.

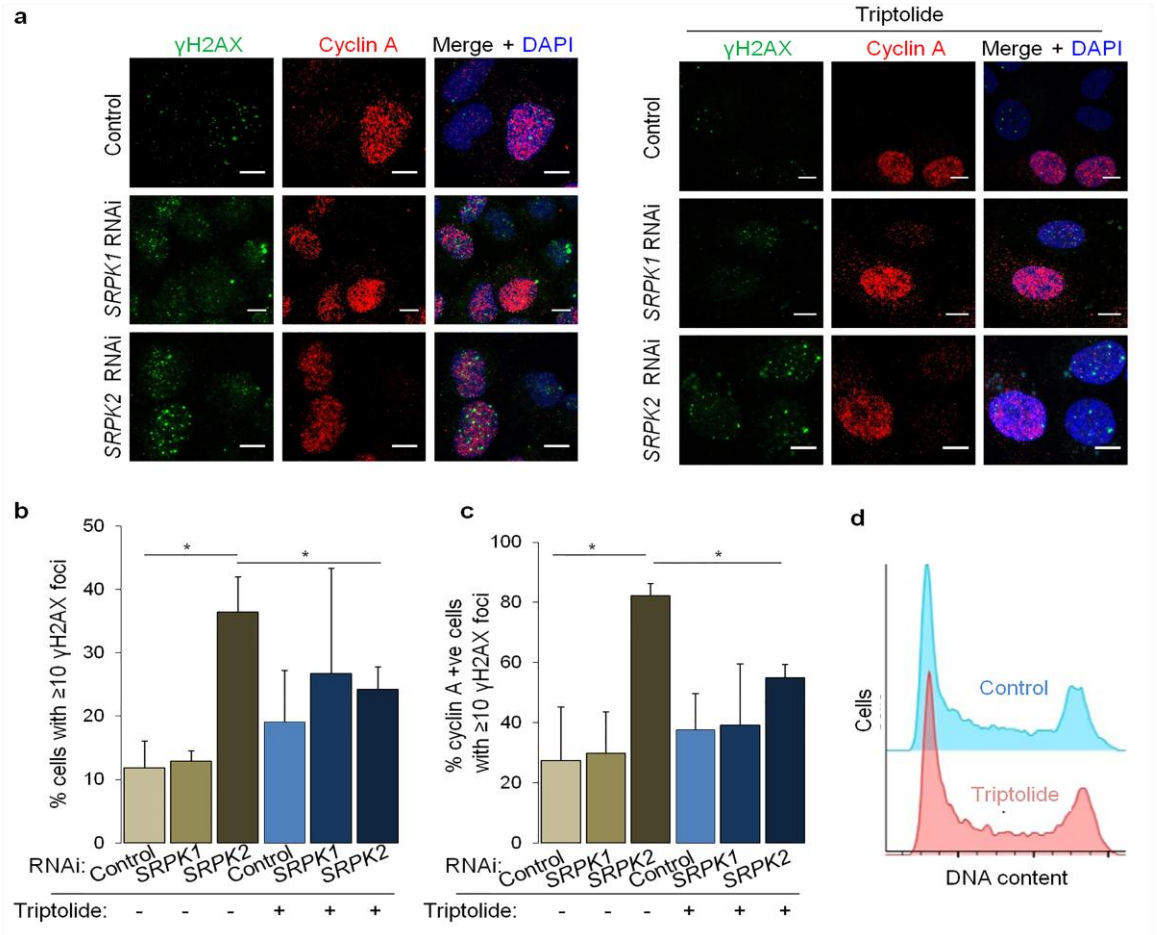


Figure 10. Transcription inhibition prevents the genome instability in *SRPK2*-depleted cells. a) γH2AX foci (green) and cyclin A staining (red) in control, *SRPK1* and *SRPK2* RNAi depleted U-2OS cells with and without triptolide treatment (to inhibit transcription). Scale bars: 10μm. **b)** Means and standard deviations of the percentage of cells with 10 or more γH2AX foci are plotted. At least 100 cells from three independent experiments were scored. **c)** same as in **(b)** but for cyclin A-positive cells only. All statistical significance was determined using two-tailed Student's t-test *p<0.05. **d)** Cell cycle progression of control and triptolide treated U-2OS cells were obtained by flow cytometry analysis of propidium iodide staining. Data shown are from one representative experiment.

After confirming the association of SRPKs with chromatin, we proceeded to perform chromatin immunoprecipitation followed by deep sequencing of the isolated DNA (ChIP-seq) to identify their whole-genome occupancy. We followed the ENCODE guidelines (Landt and Marinov, 2012) to test the specificity of SRPK1 and SRPK2 antibodies (**Fig.11a**) and fragmenting DNA to ~100-300 bp (**Fig.11b**). The SRPK1 and SRPK2 distribution along selected individual genes (**Fig.11c)** was validated by ChIP-qPCR (**Fig.11d,e**) with the same antibodies used for ChIP-seq.

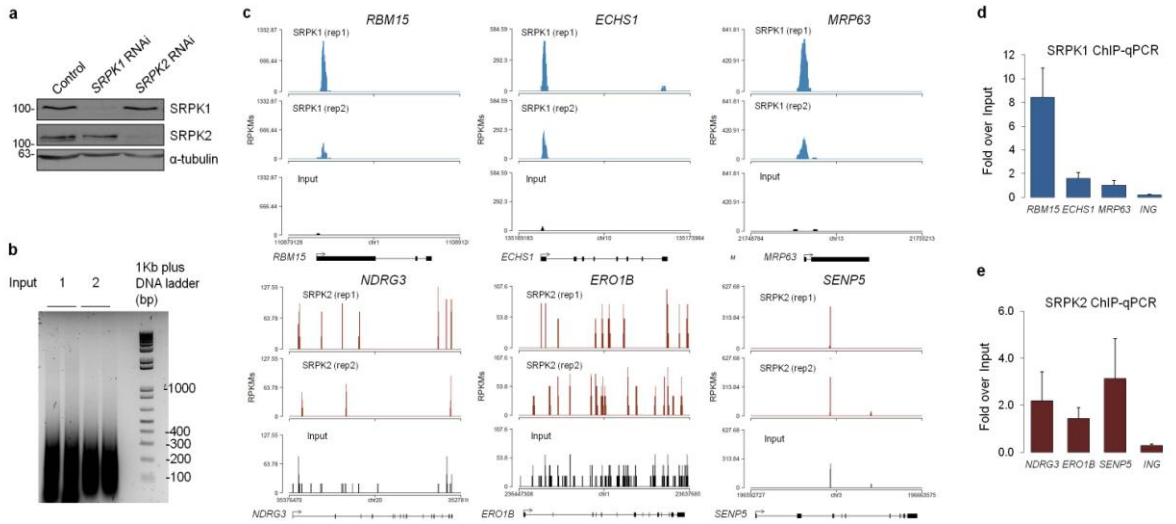


Figure 11. SRPK1 and SRPK2 ChIP-seq validation. a) Immunoblots showing SRPK1 and SRPK2 levels in transiently transfected with *SRPK1* and *SRPK2* siRNAs. α-tubulin served as loading control. **b)** Input DNA samples collected for ChIP-seq analysis were run on 1% agarose gel. Note that inputs from two biological replicates (1 and 2, including independent technical replicates) were run on gel to show that most of the DNA is about 100-300 bp fragments. 1 Kb plus DNA ladder are shown on the left in base pairs. **c)** Individual profiles of SRPK1 (blue) and SRPK2 (red) with respective duplicates at candidate gene loci. ChIP analysis of SRPK1 **(d)** and SRPK2 **(e)**, data normalized against respective input samples.

The genome-wide chromatin distribution of both SRPKs revealed that they nucleate preferentially in protein coding genes (**Fig.12a**). While SRPK1 is enriched at the promoter region, SRPK2 is distributed more evenly across the entire gene. Both kinases were preferentially detected at longer genes, but gene expression levels or GC content did not influence their chromatin binding (**Fig.12b,c**). However, we observed a strand asymmetry in the distribution of G and C residues (GC skew) in the vicinity of SRPK peaks (**Fig.12c**). Since GC skew is a predictor of R-loop formation (Ginno et al., 2013, 2012; Sanz et al.,

2016) we wondered whether aberrant accumulation of R-loops drives the genomic instability observed in *SRPK2*- depleted cells.

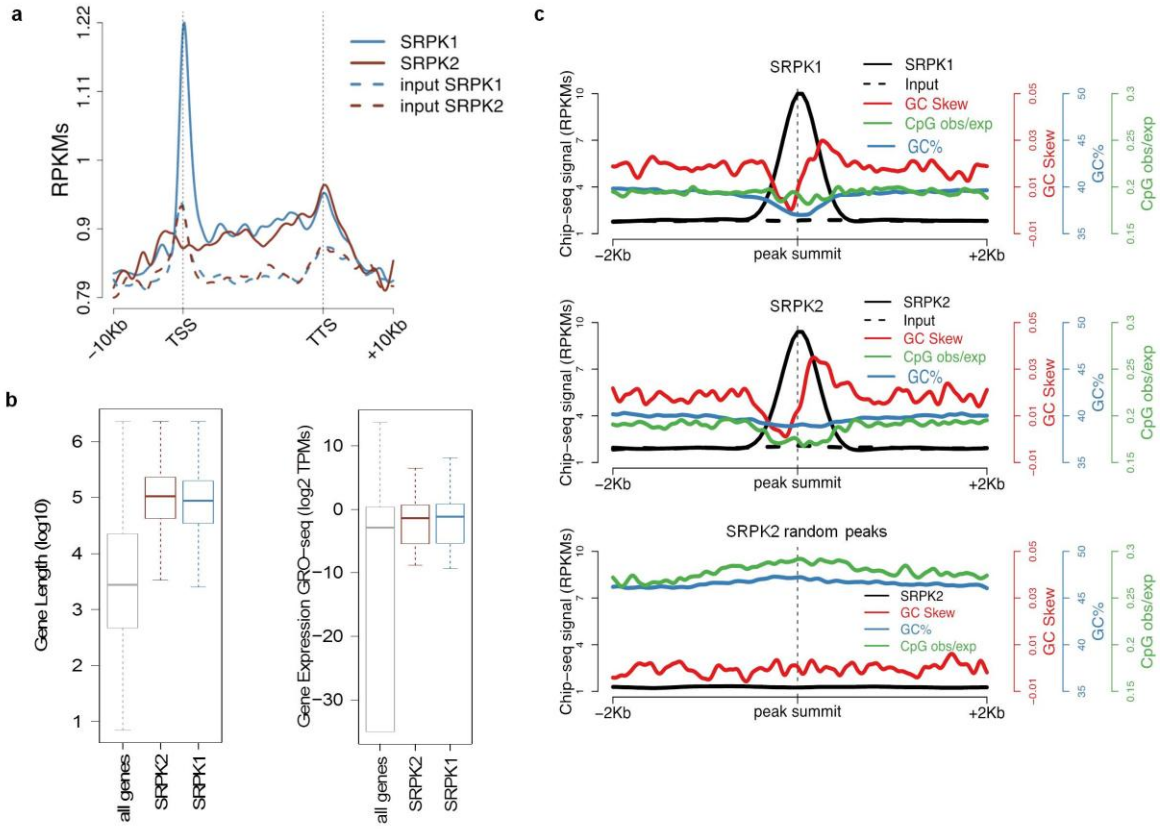


Figure 12. SRPK1 and SRPK2 occupy intragenic chromatin regions that are prone to R-loop formation. a) Metagene analysis of SRPK1 and SRPK2 genome-wide average profiles. The average ChIP-seq signal (RPKMs) of two biological replicates is shown for each kinase. The gene body region was scaled to 100 equally sized bins and ± 10 Kb gene-flanking regions were averaged in 250 bp windows. **b)** Gene Length (log10) and expression from GRO-seq (log2 TPMs) for all genes (gray); genes with SRPK2 (red) and SRPK1 (blue) enriched regions. **c)** Metaprofiles of SRPK1, SPRK2 enriched regions and random peaks dataset showing: read density (RPKMs, black solid line for ChIP and dashed line for input); GC-skew (red); CpG island (green); GC% (blue). Features were aggregated for 50 bp sliding windows of 1 bp step size.

2.3 SRPK2 prevents RNA-mediated genome instability

To test the contribution of R-loops in triggering the genome instability observed in *SRPK2*-depleted cells, we expressed RNase H1 exogenously in U2-OS cells depleted of each of the two kinases. Exogenenous expression of RNase H1 is a classical way to suppress R-loops in cells (Bhatia et al., 2014; Paulsen et al., 2014; Skourti-Stathaki et al., 2011). This experiment revealed that RNase H1 overexpression rescued the DNA damage phenotype obtained upon *SRPK2*-depletion as revealed by decreased nuclear γH2AX foci in RNase H1-positive cells (**Fig.13a**). RNase H1 overexpression also reduced the total cellular γH2AX levels upon depletion of *SRPK2*, as revealed by western blot (**Fig.13b**). Additionally, ectopic expression of RNase H1 does not affect cell cycle profile within the time frame of our experimental setup indicating that the effect on DNA damage is not the result of a deregulated cell cycle progression (**Fig. 13c**)

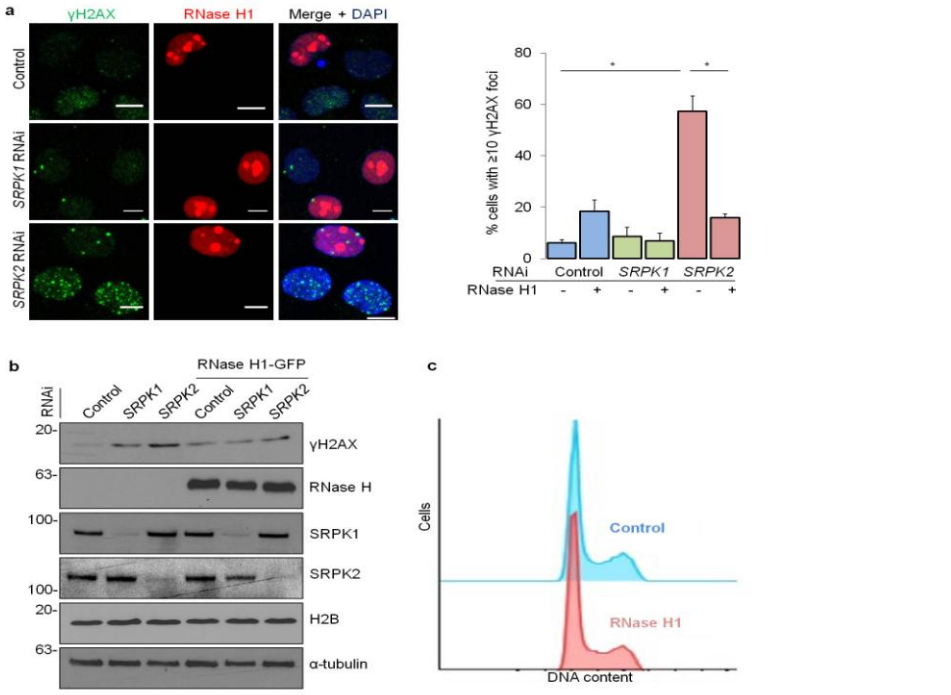


Figure 13. R-loops are the major source of genome instability in *SRPK2*-depleted cells. a) γH2AX foci (green) in control, *SRPK1* and *SRPK2* RNAi-depleted cells transiently transfected with an *RNaseH1-mCherry* expression plasmid. Means and standard deviations of the percentage of cells with ≥10 γH2AX foci are plotted on the right-hand side. Note that the apparent increase in the % of control cells with γH2AX foci upon RNase H1 overexpression (red) is not statistically significant. Data are from a minimum of 100 cells scored in four independent experiments. Scale bar: 10 μm. All statistical significance was determined using two-tailed Student's t-test. *p<0.05 **b)** Immunoblots showing γH2AX, RNase H1 (α-GFP), SRPK1 and SRPK2 levels in *SRPK1* and *SRPK2* RNAi-depleted U-2OS cells that were transiently transfected with *RNaseH1-GFP*. Histone H2B and α-tubulin served as loading controls. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments. **c)** Cell cycle progression of control and *RNaseH1-GFP* expressed U-2OS cells was obtained by flow cytometry analysis of propidium iodide staining. Data are from one representative experiment of a total of three independent experiments performed.

To directly inspect if R-loops accumulate in *SRPK2*-depleted cells we used the S9.6 antibody, which specifically detects RNA-DNA hybrids in a sequence independent manner (Boguslawski et al., 1986), in immunofluorescence experiments. We observed a robust accumulation of nuclear R-loop levels in cells devoid of *SRPK2* as revealed by the increased nucleoplasmic (non-nucleolar) signal obtained (**Fig.14**). We also observed a strong R-loop signal in nucleoli and in the cytoplasm, which is most likely the result of ribosomal DNA transcription and mitochondrial DNA replication, respectively (Brown et al., 2008; El Hage et al., 2010). **Altogether these data implicate SRPK2 in a novel mechanism that is necessary for co-transcriptional R-loops suppression and in preventing RNA-mediated genomic instability.**

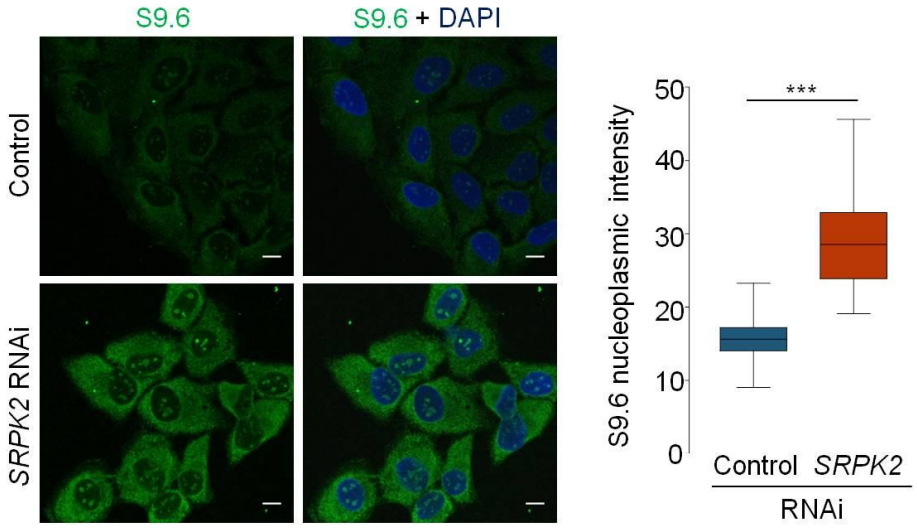


Figure 14. R-loops accumulate in *SRPK2*-depleted cells. Immunofluorescence analysis of R-loops obtained with the S9.6 antibody in control and *SRPK2* RNAi-depleted cells. The nucleoplasmic intensity of the S9.6 staining (green) is plotted on the right-hand side. At least 100 cells from three independent experiments were scored. Scale bar: 10 μm. Statistical significance was determined using Mann-whitney test. ***p<0.0005

And they call to let you know your friend is dead in a box,

The crows have the tools to get the meat out of the box,

Scientific, ritualistic, headstone cold foxes still rot.

Aesop Rock, Kimya Dawson. Crows 1

Chapter 3.
The DEAD box helicase DDX23 suppresses RNA-mediated genome instability

Highlights

- DDX23 (DEAD box helicase 23) is a specific substrate of SRPK2.
- Phosphomimetic form of DDX23 rescues DNA damage in *SRPK2*-depleted cells.
- DDX23 helicase suppresses RNA-mediated genome instability.

3.1 SRSF1 does not rescue DNA damage in cells lacking SRPK2

Based on the data discussed in **Chapter 2**, it is apparent that SRPK2 is necessary to regulate cellular R-loop levels. We then sought to investigate the molecular mechanisms of this regulatory activity. We first searched the literature for substrates that SRPK2 phosphorylates. Amongst these substrates, SRSF1 emerged as a good candidate, because of its already known role in suppressing R-loop-driven DNA damage (Li and Manley, 2005; Paulsen et al., 2014). SRSF1 is a canonical splicing protein that belongs to the family of SR proteins. Although not part of the core spliceosome, SRSF1 plays a crucial role in pre-mRNA splicing (Sapra et al., 2009). Additionally, SRSF1 prevents R-loop formation and the resulting DNA damage, possibly by preventing the annealing of the nascent RNA with the template DNA (Li and Manley, 2005). We therefore investigated whether the lack of SRSF1 phosphorylation might be responsible for the genomic instability observed in *SRPK2*-depleted cells. To test this hypothesis, we ectopically expressed wild-type SRSF1 (SRSF1-WT) or a mutant version of SRSF1 (SRSF1-PhosM) with serine-to-aspartate substitutions that mimic the phosphorylated protein in *SRPK1*- and *SRPK2*-depleted cells (Cazalla et al., 2002; Misteli et al., 1998). Both SRSF1-WT and SRSF1-PhosM exhibited a typical speckles-like pattern similar to the endogenous SRSF1 (**Fig.15a**). However, contrary to our hypothesis, ectopic expression of each SRSF1 proteins failed to rescue the DNA damage phenotype of *SRPK2*-depleted cells (**Fig.15b**). These data suggest that lack of SRSF1 phosphorylation is not the cause of the genomic instability observed in *SRPK2*-depleted cells.

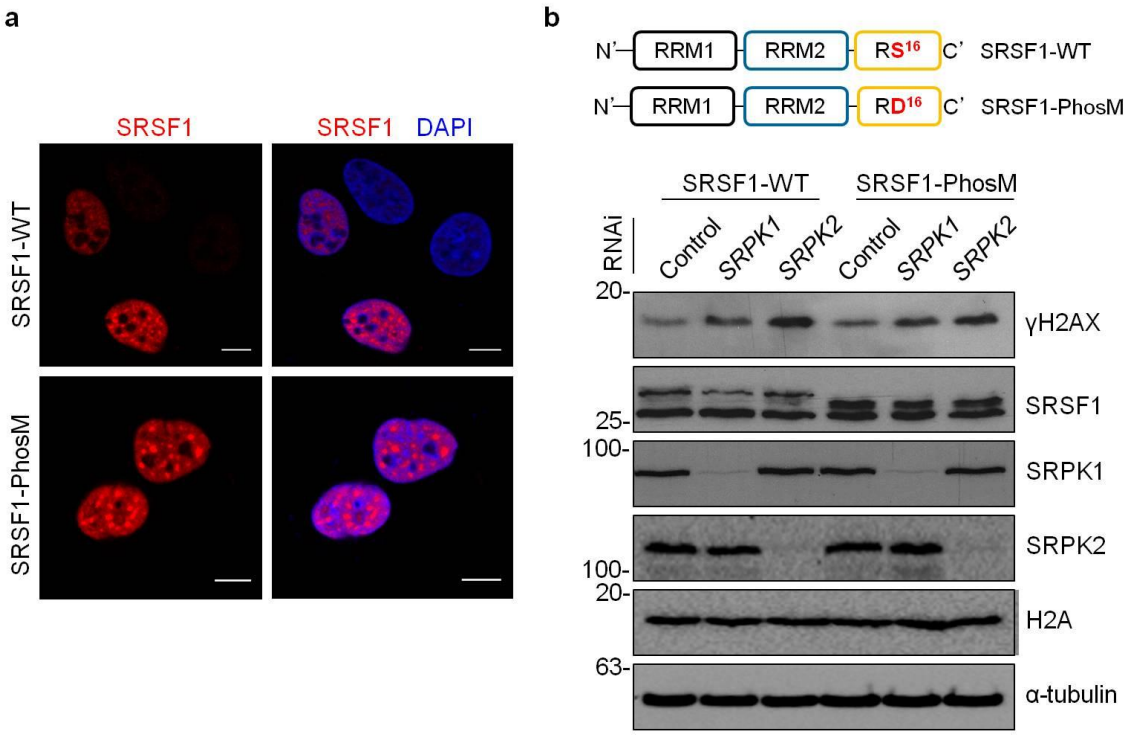


Figure 15. SRSF1 is not sufficient to rescue DNA damage in SRPK2 depleted cells. a) Immunofluorescence analysis of SRSF1-WT and SRSF1-PhosM in U-2OS cells. Scale bar: 10 µm. **b)** Immunoblots showing γH2AX, SRSF1 SRPK1 and SRPK2 levels in *SRPK1* and *SRPK2* RNAi-depleted U-2OS cells that were transiently transfected with: *SRSF1-WT* or *SRSF1-PhosM*. Histone H2A and α-tubulin served as loading controls. Data are representative of three independent experiments.

3.2 DEAD box helicases

The failure of SRSF1 to resolve DNA damage in *SRPK2*-depleted cells instigated us to look for alternative SRPK2 specific substrates. Our search revealed DDX23, a specific SRPK2 substrate that is not phosphorylated by SRPK1 (Mathew et al., 2008). DDX23 is a RNA helicase that belongs to the DEAD box class.

RNA helicases lie at the heart of RNA processing from transcription to translation and decay. They are present in different cellular compartments (like cytosol, mitochondria, and nucleus) because of their diverse functions in different steps of RNA maturation. Among the different helicase families, DEAD box proteins form the largest group. As the name

suggests, they have a conserved DEAD box motif that has aspartate(D)-glutamate(E)-alanine(A)-aspartate(D) aminoacids (Jankowsky, 2011; Jarmoskaite and Russell, 2014). They orchestrate the functions of super molecular machines like the spliceosome by structurally remodeling RNA and protein complexes (Will and Lührmann, 2011). Mutations in DEAD box proteins have been attributed to a wide variety of diseases, including cancer (Yin et al., 2015).

3.2.1 Structure of DEAD box helicases

A typical DEAD box protein contains 12 highly conserved sequence motifs. The 'DEAD box' is part of the motif II (**Fig.16a**). These evolutionary conserved enzymes use energy from ATP to remodel RNA and RNPs. ATP and RNA binding sites reside between the two helicase domains on either side. The two helicase domains form a closed cleft, to productively bind and hydrolyse ATP. Multiple amino acids are responsible for interactions with ATP and RNA independently, making it difficult to inactivate DEAD box proteins (**Fig.16b**) (Rocak and Linder, 2004). DEAD box proteins bind to the sugar phosphate backbone of RNA (~5 nucleotides), bending the RNA in a conformation characteristic only to DEAD box group of helicases. The helicase domain attains a more orderly conformation in the presence of either RNA or ATP. However, it is still unclear how the RNA and ATP binding sites signal each other (Linder and Jankowsky, 2011). Regulatory elements reside in either N-terminal and/or C-terminal domains that flank the helicase domain. Most of the data on DEAD box proteins arise from peptides that form the helicase domain, giving very little information on these regulatory elements. For instance, human DDX23 (unlike yeast Prp28), retains a RS-domain in its N-terminal regulatory domain that is specifically phosphorylated by SRPK2 (Mathew et al., 2008; Muhlmann et al., 2014).

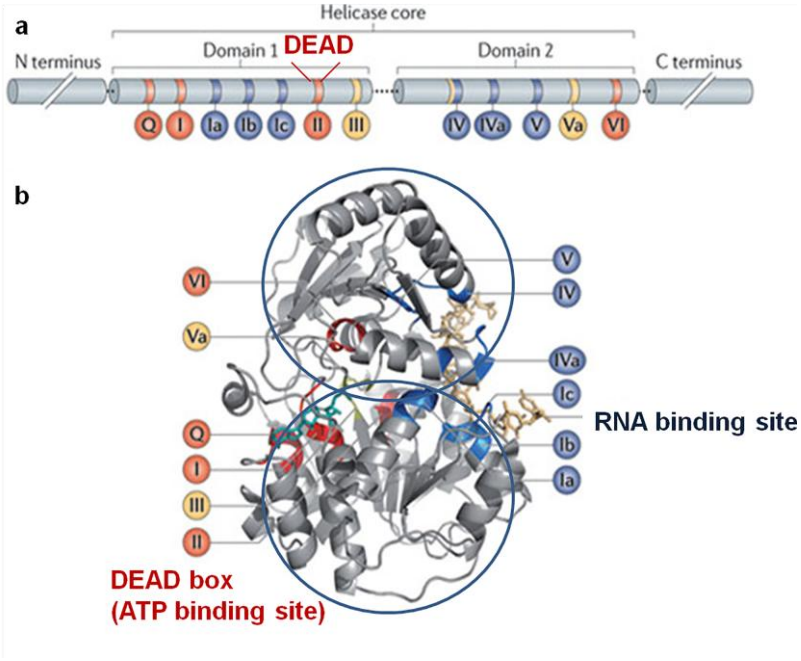


Figure 16. Schematic view of DEAD box helicases. a) Canonical DEAD box helicase with conserved sequence motifs (Q, I, Ia, Ib, Ic, II, III, IV, Iva, V, Va, VI). **b)** Structure of the helicase core domains of the DEAD box protein Vasa from *D. melanogaster*. Note the ATP and RNA binding sites on either side of helicase core.

3.2.2 Mechanism of action of DEAD box helicases

One of the major differences between DEAD box helicases and other DNA/RNA helicases is the lack of unwinding polarity. The lack of unwinding polarity allows the DEAD box helicase to unwind RNA/DNA in any direction (5'→3' or 3'→5') (Jankowsky, 2011) DEAD box helicases can also load onto RNA at any of its extremities and in a manner that depends on the RNA secondary structure (e.g., CYT19 from *Neurospora crassa* and bacterial DbpA) (Tijerina et al., 2006). DEAD box helicases catalyze both ATP-dependent and independent functions that promote complex RNA rearrangements (Linder and Jankowsky, 2011). For example, nucleolar DDX21 displays intrinsic ATP-independent activity (Calo et al., 2014). The RNA remodeling/helicase activity discrepancies exhibited by DEAD box proteins in an *in vitro* or *in vivo* setting suggest the requirement of co-factors and/or other cell signaling pathways for their activity. For example, DEAD box

proteins like Ded1 and Mss116 can act as 'strand annealers' independently of ATP (Jarmoskaite and Russell, 2014; Linder and Jankowsky, 2011). Once bound to RNA, DEAD box helicases start strand separation taking the energy from ATP hydrolysis. The unwinding rate constant is indirectly proportional to length and stability of duplex and presence of ADP. Depending on the processivity and ATP hydrolysis, the helicase reaction can be productive (resulting in the unwinding of duplex) or futile (re-annealing of duplex after the helicase dissociation) (Jankowsky, 2011; Linder and Jankowsky, 2011).

3.2.3 DDX23

Although small nuclear RNA (snRNA) forms the catalytic core of the spliceosome, various accessory proteins lie at the basis of the dynamic molecular rearrangements that dictate pre-mRNA splicing. From early to late splicing factors, there is a dynamic interplay between different splicing factors and DExD/H box helicases (DEAD box, DEAH and the SKI families of proteins are all referred to as DExD/H box proteins) play central role in such molecular rearrangements (Wahl et al., 2009). DDX23 (also termed hPrp28/U5-100K, yeast Prp28) and SNRNP200 (yeast Brr2) are involved in the transitions from a pre-catalytic spliceosome to an activated spliceosome. Studies in yeast, revealed that Prp28 drives the displacement of the U1 snRNP from the 5' splice site and proof-reads the splice sites (**Fig.17a**). Further, mutagenesis studies demonstrated the requirement of both the Prp28 ATPase and helicase activities for pre-mRNA splicing (Chang et al., 1997). However, ATPase-independent Prp28 activity was also reported to occur during splicing (Price et al., 2014). This variety of functions suggests that a complex molecular network regulates Prp28 activity, through mechanisms that might involve post-translational modifications such as phosphorylation. In human cells, DDX23 which is a part of U5snRNP complex contains around 230 N-terminal residues that are rich in arginine/serine, arginine/glutamate and arginine/aspartate dipeptide motifs and therefore have been annotated as an RS-like domain (**Fig. 17b**) (Teigelkamp et al., 1997). The DDX23 RS domain is phosphorylated specifically by SRPK2 (Mathew et al., 2008; Wahl et al., 2009). This phosphorylation drives a dynamic switch required for the association of DDX23 with the U4/U6.U5 tri-snRNP (Mathew et al., 2008; Xiang et al., 2013). In agreement, *SRPK2* knockdown results in hypophosphorylation of DDX23 and destabilizes its association with the U5 snRNP (Boesler et al., 2016; Mathew et al., 2008).

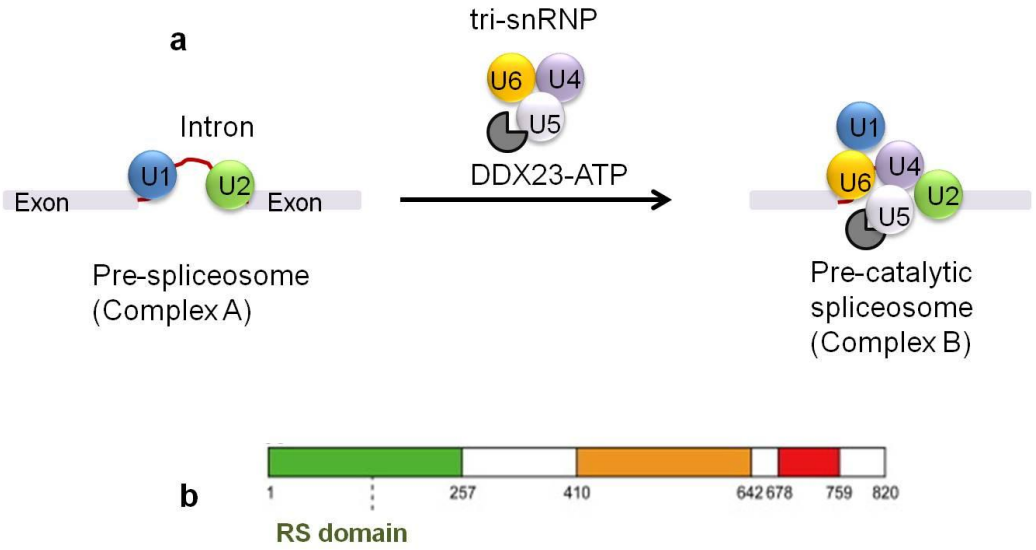


Figure 17. Role of DDX23 in pre-mRNA splicing. a) The stepwise interaction of the core-spliceosomal proteins (colored circles) in the removal of an intron (red line) from a pre-mRNA containing two exons (gray boxes). Transition from complex A to complex B has been shown here, where splicing reaction starts with binding of U1 snRNP to 5' splice site and U2 snRNP to 3'splice site. The tri-snRNP compex (U4/U6.U5 snRNPs) consisting of active phosphorylated DDX23 (dark gray) aligns on complex A to form complex B with the displacement of U1 snRNP from 5' splice site. Yeast Prp28 acts at a later stage during spliceosome activation (the B complex to B* complex transition). **b)** Disordered domain (1–257, green), DEAD box domain (410–642, orange), and HELICc domain (678–759, red) of DDX23 with RS domain (1–138). Note that, yeast Prp28 lacks N-terminal RS domain. Adapted from Linder and Jankowsky, 2011; Xiang et al., 2013.

3.3 A phosphomimetic DDX23 rescues the genome integrity in *SRPK2*-depleted cells

To address the role of DDX23 in the SRPK2-associated genomic instability phenotype, we made two constructs of DDX23: 1) a wild type DDX23 (DDX23-WT) and 2) a phosphomimetic version of DDX23 (DDX23-PhosM). To make DDX23-PhosM, we replaced 9 serine residues in the RS domain with 9 aspartates. The choice of the serines that were mutated was made based on previously reported data (Teigelkamp et al., 1997). Both the DDX23-WT and DDX23-PhosM were tagged with GFP. Analysis of their sub-

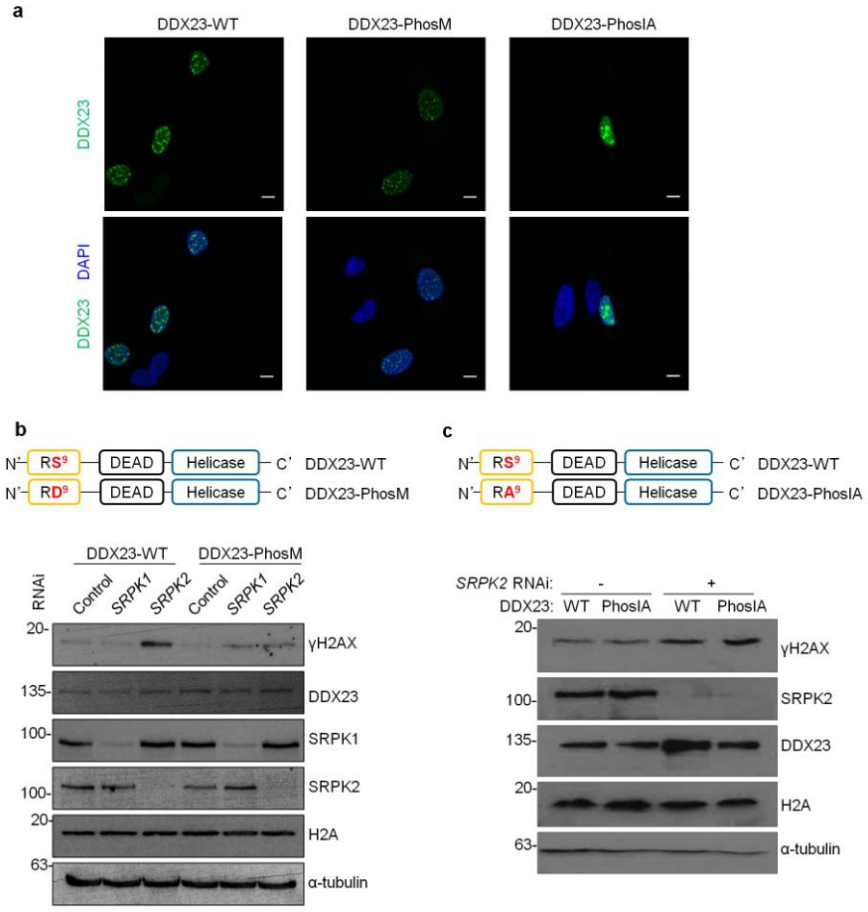


Figure 18. A phosphomimetic version of DDX23 rescues the genome integrity in *SRPK2*-depleted cells. a) Immunofluorescence analysis of DDX23-WT, DDX23-PhosM, and DDX23-PhosIA in U-2OS cells. Scale bar: 10 μm. **b)** Immunoblots showing γH2AX, DDX23, SRPK1 and SRPK2 levels in *SRPK1* and *SRPK2* RNAi-depleted U-2OS cells that were transiently transfected with DDX23-WT or DDX23-PhosM. **c)** Immunoblots showing γH2AX, SRPK2 and DDX23 in U-2OS cells following RNAi-depletion of *SRPK2* and transient transfection of DDX23-WT or DDX23-PhosIA as detailed in the figure. Molecular weight markers (KDa) are shown on the left. α-tubulin and histone H2A levels served as loading controls. Data are representative of three independent experiments.

cellular localization upon overexpression in U2-OS cells revealed a nuclear speckles-like pattern typical of other RNA-processing proteins (**Fig.18a**). We then ectopically expressed both DDX23 proteins in *SRPK1* or *SRPK2*-depleted cells to assess their effect on the DNA damage (evaluated by measuring γH2AX levels). As expected, knockdown of *SRPK2* but not *SRPK1* induced γH2AX (**Fig.18b**). Notably, ectopic expression of DDX23-PhosM, but not of DDX23-WT, was able to completely abolish the DNA damage induced by the *SRPK2* knockdown (**Fig.18b**). We then made a third version of DDX23 where the same 9 serine residues of the RS domain were mutated to 9 alanines. Mutation of serine to alanine blocks DDX23 phosphorylation rendering the protein phospho-inactive (DDX23-PhosIA). This protein also localized to nuclear speckles and was also detected in nucleoli (**Fig.18a**). Notably, ectopic expression of DDX23-PhosIA failed to reduce γH2AX levels in *SRPK2*-depleted cells (**Fig.18c**). These data suggest that DDX23 phosphorylation is necessary and sufficient to rescue the genome integrity in *SRPK2*-depleted cells.

We then asked if the helicase domain of DDX23 is required to suppress the DNA damage in *SRPK2*-depleted cells. To this end, we constructed a truncated version of DDX23-PhosM, which we named DDX23-PhosM-ΔHel. This protein lacks most of the helicase domain while maintaining the serine-to-aspartate substitutions of the phosphomimetic version of DDX23. Again, we could detect DDX23-PhosM-ΔHel in nuclear speckles and associated with chromatin (**Fig.19a,b**). We then overexpressed the three versions of DDX23 (WT, PhosM and PhosM-ΔHel) in control and *SRPK2* RNAi U2-OS cells. In agreement with our previous results, DDX23-PhosM but not DDX23-WT was able to completely rescue the γH2AX levels upon *SRPK2* depletion. Notably, the phosphomimetic DDX23 lacking the helicase domain (DDX23-PhosM-ΔHel) failed to reduce the DNA damage phenotype of *SRPK2*-depleted cells as demonstrated by western blot and immunofluorescence (**Fig.19c,d**). Altogether, these data show that phosphorylation of DDX23 and its helicase activity are absolutely necessary to counteract the DNA damage phenotype obtained upon *SRPK2* depletion.

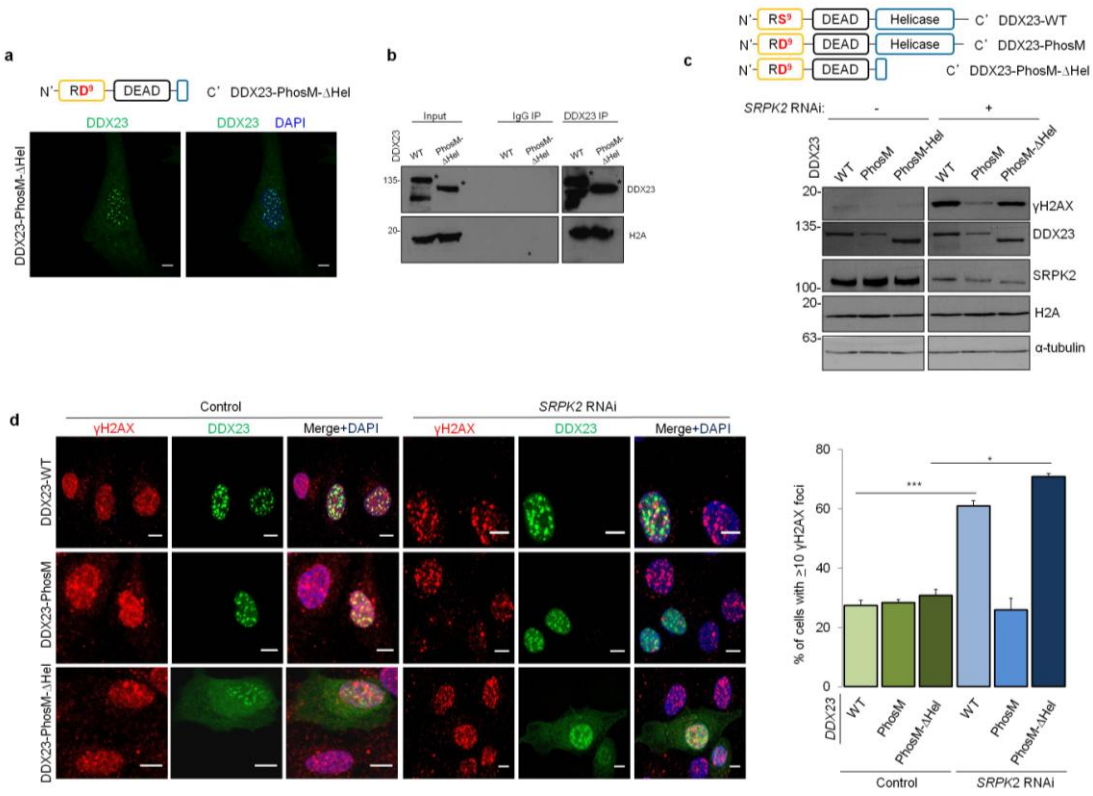


Figure 19. Helicase activity of DDX23 is necessary to regulate RNA-mediated genome instability. a) Immunofluorescence analysis of DDX23-PhosM-ΔHel in U-2OS cells. Scale bar: 10 µm. **b)** Co-Immunoprecipitation of wild type (WT) and helicase-dead DDX23 mutants (PhosM-ΔHel) with histone H2A in U-2OS cells. DDX23-purified complexes were resolved by SDS-PAGE and blotted with antibodies against GFP (to detect DDX23) or histone H2A. The Input lane represents total cell lysates and IgG IP denotes the negative control immunoprecipitation obtained with an isotype-matched antibody. The asterisks indicate the bands corresponding to GFP-tagged DDX23 proteins. Molecular weight markers (KDa) are shown on the left. **c)** γH2AX, DDX23 and SRPK2 levels in *SRPK2* RNAi-depleted U-2OS cells that were transiently transfected with DDX23-WT, DDX23-PhosM or DDX23-PhosM-ΔHel. Histone H2A and α-tubulin served as loading controls. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments. **d)** γH2AX foci and DDX23-GFP in control and *SRPK2* RNAi-depleted U-2OS cells upon transient transfection with either DDX23-WT, DDX23-PhosM or DDX23-PhosM-ΔHel. Means and standard deviations of the percentage of cells with ≥ 10 γH2AX foci are plotted on the right-hand side. Data are from a minimum of 100 cells scored in three independent experiments. Scale bar: 10 µm. Statistical significance was determined using two-tailed Student's t-test. *p<0.05, ***p<0.0005

We next sought to investigate if the role of DDX23 in suppressing R-loops requires a functional U5 snRNP. To this end, we performed RNAi against *PRP8*, a U5 snRNP scaffolding protein PRP8, which is at the core of U5 snRNP assembly and plays key roles during the catalytic activation of the spliceosome (Agafonov et al., 2016; Grainger and Beggs, 2005; Ritchie et al., 2008). Depletion of *PRP8*, did not induce any DNA damage (**Fig.20a)** and did not affect the ability of DDX23-PhosM to suppress DNA damage in *SRPK2*-depleted cells (**Fig.20b-d)**. Similar data were obtained upon depletion of *PRP6*,

another U5 snRNP component that is necessary for spliceosome activity (Makarov et al., 2000) (**Fig.21)**. Altogether, these results show that phosphorylation of DDX23 restores the genome stability in *SRPK2*-deficient cells through a process that does not require a functional U5 snRNP or spliceosome activity.

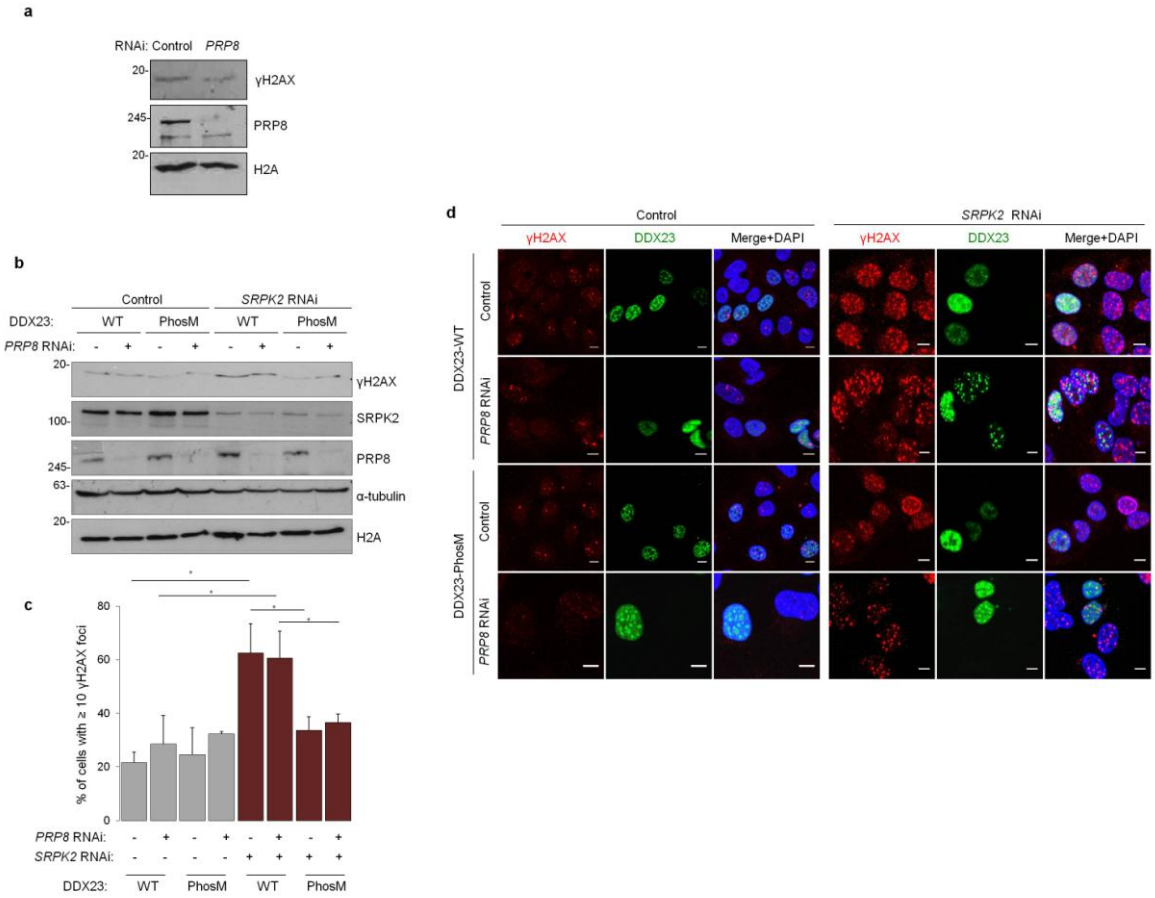


Figure 20. PRP8 is not necessary for the suppression of DNA damage by DDX23. a) γH2AX and PRP8 levels in *PRP8* RNAi-depleted U-2OS cells. **b)** γH2AX, SRPK2 and PRP8 levels in U-2OS cells following RNAi-depletion of *SRPK2* and *PRP8* and transiently transfected with *DDX23-WT* or *DDX23-PhosM*, as detailed in the figure. Molecular weight markers (KDa) are shown on the left. α-tubulin and histone H2A levels served as loading controls. Data are representative of three independent experiments. **c)** and **d)** γH2AX foci and DDX23 staining in control, *PRP8*, *SRPK2* RNAi depleted U-2OS cells upon transient transfection with DDX23-WT, DDX23-PhosM. Means and standard deviations of the percentage of cells with 10 or more γH2AX foci are plotted. Data are from a minimum of 100 cells scored in three independent experiments. Scale bars: 10μm. Satistical significance was determined using two-tailed Student's t-test. *p<0.05

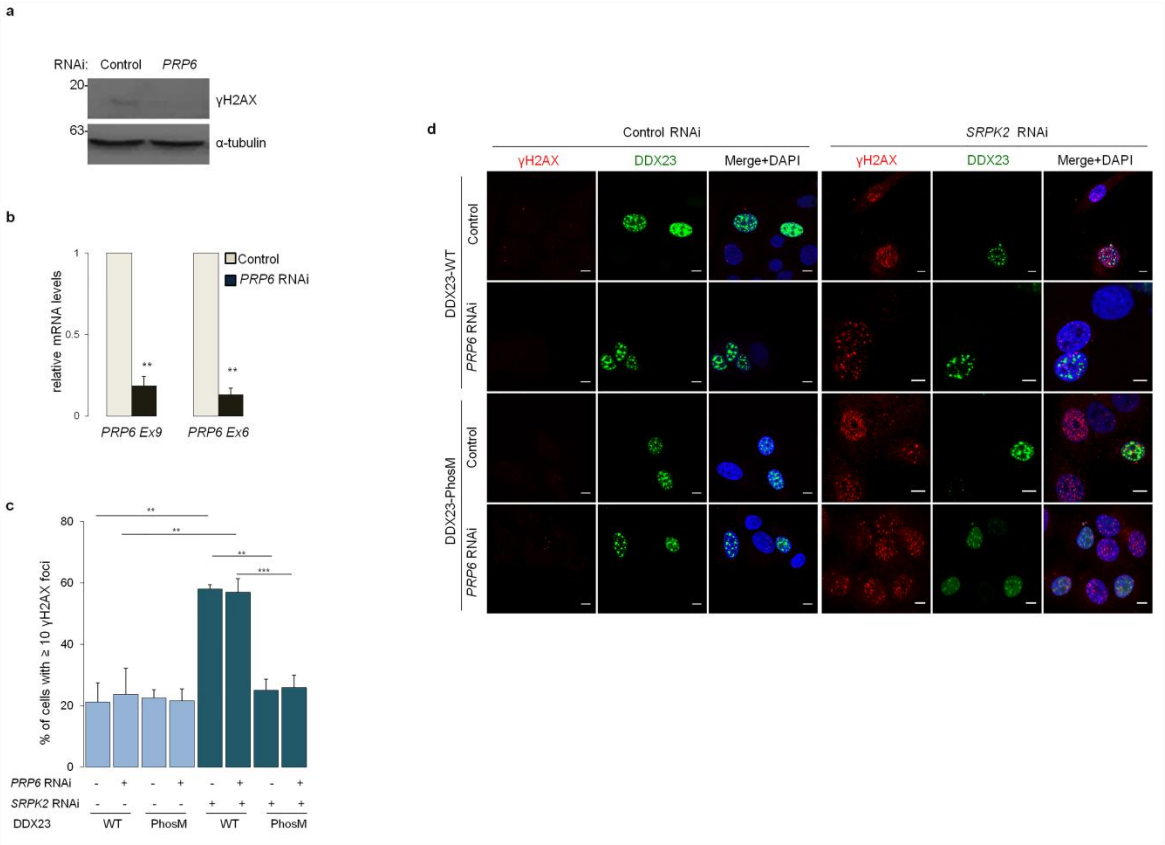


Figure 21. PRP6 is not necessary for the suppression of DNA damage by DDX23. a) γH2AX in *PRP6* RNAi-depleted U-2OS cells. Molecular weight markers (KDa) are shown on the left. α-tubulin levels served as loading controls. Data are representative of three independent experiments.. **b)** *PRP6* mRNA levels in control and *PRP*6 RNAi treated U2-OS cells as assessed by primers against exon 9 and 6 of human *PRP6* locus. Data was normalized against values obtained in control (i.e. without *PRP6* knockdown) cells. Means and standard deviations are from a minimum of three independent qPCR experiments. Statistical significance was determined using two-tailed Student's t-test. **p<0.005, ***p<0.0005 **c)** and **d)** γH2AX foci and DDX23 staining in control, *PRP8, SRPK2* RNAi depleted U-2OS cells upon transient transfection with DDX23-WT, DDX23-PhosM. Means and standard deviations of the percentage of cells with 10 or more γH2AX foci are plotted. Data are from a minimum of 100 cells scored in three independent experiments. Scale bars: 10μm. Statistical significance was determined using two-tailed Student's t-test. **p<0.005, ***p<0.0005

3.4 Loss of DDX23 drives R-loop-dependent genome instability

To further explore the role of DDX23 in maintaining genome stability, we measured cellular R-loop levels upon transient *DDX23* knockdown. *DDX23* was depleted using an independent pool of siRNA. A control RNAi included a pool of siRNA directed against firefly luciferase gene. In agreement with a role in R-loop associated genome stability, transient depletion of *DDX23* resulted in a strong accumulation of nuclear R-loops (**Fig.22a**). Notably, there was also a prominent increase in total γH2AX levels as assessed by immunoblots of protein lysates collected from control and *DDX23* RNAi treated cells (**Fig.22b**). Strikingly, ectopic expression of RNase H1 reduced nuclear γH2AX foci

observed upon depletion of *DDX23*, suggesting that R-loops are the major source of the DNA damage observed in *DDX23* deficient cells (**Fig.22c**).

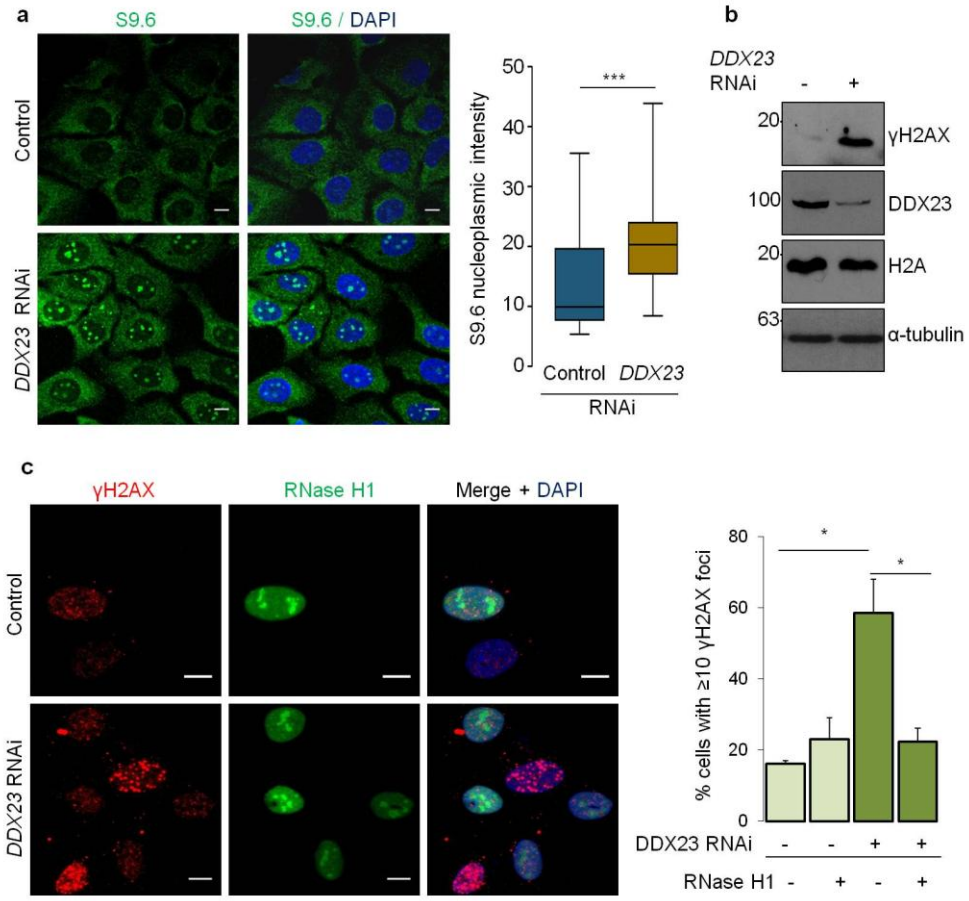


Figure 22. Loss of *DDX23* drives RNA-dependent genomic instability. a) Immunofluorescence analysis of R-loops obtained with the S9.6 antibody in control and *DDX23*-depleted U-2OS cells. The nucleoplasmic intensity of the S9.6 staining is plotted on the right-hand side. At least 100 cells from three independent experiments were scored. Statistical significance was determined using Mann-whitney test. ***p<0.0005 Scale bar: 10µm. **b)** Immunoblots showing γH2AX and DDX23 levels in control and *DDX23*-depleted U-2OS cells. Histone H2A and α-tubulin served as loading controls. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments. **c)** γH2AX foci in control and *DDX23*-depleted U-2OS cells transiently transfected with an *RNase H1-GFP* expression plasmid. Means and standard deviations of the percentage of cells with ≥10 γH2AX foci are plotted on the right-hand side. Data are from a minimum of 100 cells scored in three independent experiments. Scale bar: 10 µm. All statistical significance was determined using two-tailed Student's t-test. *p<0.05

We also observed phosphorylation of p53-binding protein 1 (53BP1) and Replication protein A (RPA) proteins (which are molecules involved in the repair of DNA DSBs) upon depletion of *DDX23* and *SRPK2* (**Fig.23a**). Additionally, in order to inspect the direct effect of SRPK2 and DDX23 on genome instability we also scored the number of DSB-associated chromosome aberrations (like single- and double- chromatid breaks). For this purpose, we collected metaphase chromosomes from cells depleted of *SRPK2* and *DDX23*. We observed that both *SRPK2* and *DDX23*-depleted cells accumulate more DSB-associated chromosomal aberrations than control cells exhibiting a strong genome instable phenotype (**Fig. 23b)**. Altogether, these data disclose an important new role of DDX23 in preventing R-loop-dependent genome instability.

In agreement with the view that phosphorylation of DDX23 by SRPK2 constitutes a new genome caretaker mechanism to ward-off against tumorigenic processes such as genomic instability, our pan-cancer data analysis revealed that 45% (24 out of 53) of all DDX23 point mutations found in cancer samples are significantly enriched in the RS domain (Fisher's Exact Test p-value < 0.05, **Fig. 23c**). Moreover, 83% of these mutations are predicted to be potentially deleterious variants (**Extended Data Table 2**). Within this domain, 63% of the mutations affected either an arginine (R) or a serine (S) residue linking to the fact that SRPK preferentially phosphorylate serines than threonine close to arginine residues (Ghosh and Adams, 2011). Additionally, DDX23 copy-number alterations were observed in several cancers, such as adenoid cystic carcinoma (ACC). Our analyses detected homozygous deletions of the DDX23 locus in 17% (10 out of 60) ACC samples (**Fig. 23d**). ACC from the salivary glands is characterized by perineural invasion and high rates of metastasis to distant organs like lungs, liver and bone (Spiro, 1997). Chromosomal translocations leading to formation of the *MYB-NFIB* gene fusion is a recurrent molecular feature of ACC. However, there is no successful chemotherapy available to treat ACC. Indeed, treatment of ACC is confined to surgery and radiation, while chemotherapy has been of limited palliative benefit in patients with advanced disease (Chae et al., 2015; Spiro, 1997). Therefore, the marked genomic instability that follows deletion of DDX23 may explain the recurrent high-level losses in 12q13−the genomic housing of DDX23−previously observed in ACC (Ho et al., 2013).

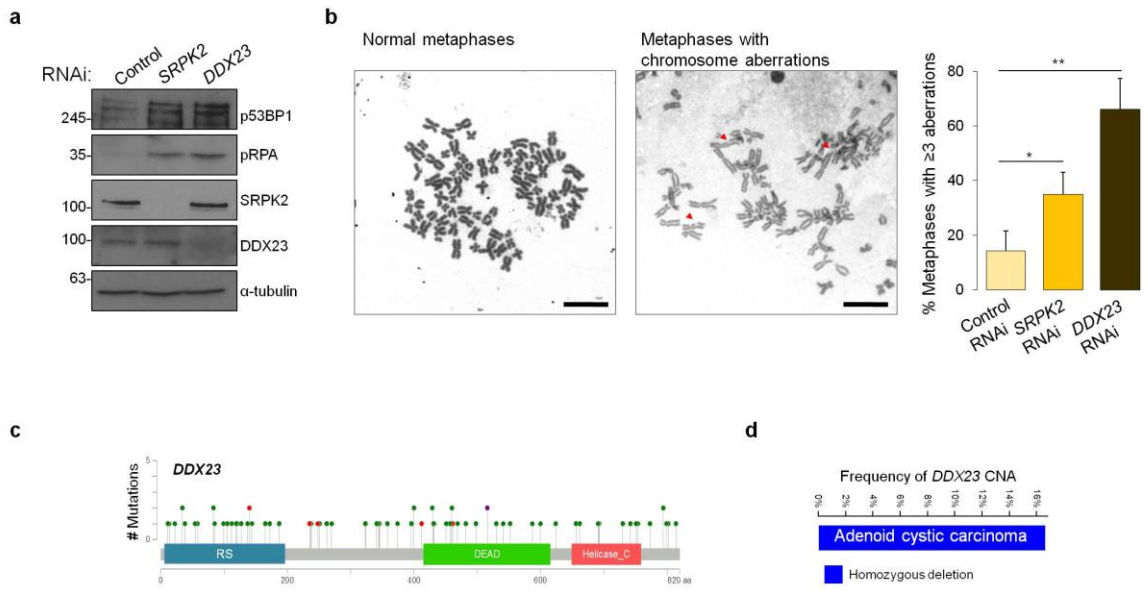


Figure 23. *DDX23* depleted cells exhibit severe genome instability phenotype. a) Immunoblots showing phosphorylated 53BP1 (p53BP1), phosphorylated RPA (pRPA), SRPK2 and DDX23 levels in SRPK2 and DDX23 RNAi-depleted HeLa cells. α-tubulin served as a loading control. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments. **b)** Typical metaphase spread of HeLa cells treated with siRNA against *SRPK2* and *DDX23* showing two different sub-populations characterized as normal-like and metaphase with ≥3 aberrations. Red arrows point to chromosomal aberrations. Percentage of metaphases with ≥3 chromosome aberrations was scored from three independent experiments with mean and standard deviations were shown on right side. At least 80 spreads were scored per condition. Scale bar: 10 μm. Statistical significance was determined using two-tailed Student's t-test. **p<0.005, *p<0.05. **c)** Schematic representation of the point-mutations found in the DDX23 locus in the cancers represented in the cBioPortal (117 studies with mutation data). The RS, DEAD and helicase domains of DDX23 are shown. **d)** Frequency of copynumber alterations (CNA) of the DDX23 locus in ACC samples. Data are from the MSKCC cBioPortal for Cancer Genomics.

In a nutshell,

1) **Phosphorylation of DDX23 is necessary and sufficient to restore the DNA damage in *SRPK2*-deficient cells through a process that requires its helicase activity.**
2) **SRPK2-DDX23 is a novel signaling axis employed by cells to suppress RNA-mediated genome instability.**
3) **R-loops are the major source of DNA damage in *DDX23*-deficient cells.**
4) **Dysfunction of DDX23 can explain genome instability observed in ACC.**

RNA Pol II, "Here is another tale about my tail, along the trail".

Anonymous

Even if you're on the right track, you'll get run over if you just sit there.

Will Rogers

Transcription: breaking up can be hard to do!

Rosonina, Kaneko, Manley

If you shut the door to all errors, truth will be shut out.

Tagore

Chapter 4. Paused RNA Pol II acts as an R-loop 'sensor-signal'

Highlights

- DDX23 nucleates at R-loops-containing loci.
- R-loops cause RNA Pol II pausing.
- Paused RNA Pol II acts as an R-loop 'sensor-signal'.

4.1 DDX23 accumulates at R-loop-containing chromatin loci

As DDX23 is necessary to maintain cellular R-loop levels, we asked if this helicase is recruited to R-loops-forming loci. To test this, we induced the formation of R-loops by depleting *THOC1*, a component of the THO/TREX complex that, if depleted, leads to aberrant accumulation of R-loops (Bhatia et al., 2014; Domínguez-Sánchez et al., 2011). First, we transiently depleted *THOC1* and *DDX23* using pooled siRNAs and assessed the level of endogenous R-loops at *APOE* and *ACTB* loci (**Fig.24a**). It was previously shown that *THOC1* depletion induces R-loop levels at these loci (Bhatia et al., 2014; Skourti-Stathaki et al., 2011) . We then used the S9.6 antibody to isolate R-loops by performing DNA-RNA immunoprecipitation (DRIP). As expected, in the absence of *THOC1*, we observed a significant increase in R-loop levels in *APOE* and *ACTB* but not in *SNRPN* (already described to lack R-loops even after *THOC1* depletion (Bhatia et al., 2014; Sanz et al., 2016) (**Fig.24b**). In agreement with the role of R-loops in inducing DNA damage, we also observed a significant increase in γH2AX levels in *APOE* and *ACTB* loci in an R-loop dependent manner as assessed by γH2AX ChIP (**Fig.24c**). Furthermore, in agreement with a role in the suppression of R-loops, *DDX23* knockdown also resulted in a prominent accumulation of these structures in *APOE* and *ACTB* (**Fig. 24b**).

We then explored if DDX23 is recruited to the chromatin regions where R-loops form upon *THOC1* knockdown. In fact, ChIP experiments showed a robust accumulation of DDX23 in R-loop-containing *APOE* and *ACTB* loci, but not in the R-loop-free *SNRPN* locus (**Fig.25a**). Importantly, suppression of R-loops by overexpression of RNase H1 significantly reduced the recruitment of DDX23 to *APOE* and *ACTB* loci in *THOC1*-depleted cells (**Fig.25a**). In contrast, this recruitment was not affected by *SRPK2* depletion (**Fig.25b**), suggesting that DDX23 phosphorylation is required for the protein activity (namely to resolve co-transcriptional R-loops) but not to its chromatin binding. These

results reveal that DDX23 accumulates at chromatin regions containing R-loops and is necessary to suppress locus-specific R-loop formation.

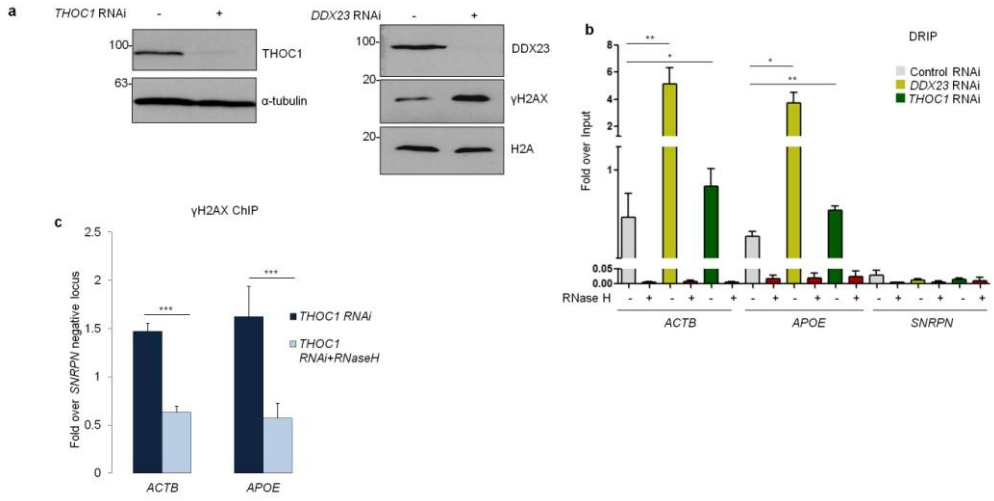


Figure 24. *THOC1* and *DDX23* depleted cells accumulate R-loops at gene loci . a) Immunoblots of THOC1, DDX23 and γH2AX in U2-OS cells depleted of THOC1 and DDX23. α-tubulin and H2A served as loading controls. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments. **b)** DNA-RNA immunoprecipitations (DRIP) in *DDX23* and *THOC1* RNAi-depleted cells is shown in the right-hand side. Data are presented as fold increase over the input sample. Means and standard deviations from three independent DRIP experiments are plotted. Statistical significance was determined using two-tailed Student's t-test. *p < 0.05 **p<0.005. **c)** ChIP analysis of γH2AX in *THOC1* RNAi-depleted cells with (+ RNase H1) or without transient transfection of RNase H1. Data was normalized against the ChIP values obtained in *SNRPN* negative locus. Means and standard deviations are from a minimum of three independent experiments. *p<0.05, **p<0.005, ***p<0.0005 (two-tailed Student's t-test).

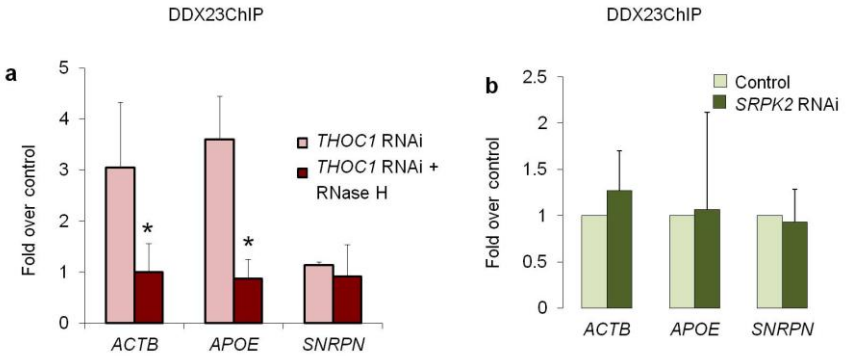


Figure 25. DDX23 nucleates at R-loop-containing sites. a) ChIP analysis of DDX23 in *THOC1* RNAi-depleted cells with (+ *RNase H1-GFP*) or without transient transfection of *RNase H1*. Data was normalized against the ChIP values obtained in control cells (i.e. without *THOC1* knockdown). Means and standard deviations from three independent DRIP experiments are plotted. Statistical significance was determined using two-tailed Student's t-test. *p < 0.05 **p<0.005. **b)** ChIP analysis of DDX23 in *SRPK2* RNAi-depleted U-2OS cells. Data are shown as fold enrichment over the ChIP values obtained in control cells. Means and standard deviations are from a minimum of three independent ChIP experiments.

4.2 RNA Pol II pausing nucleates SRPK2 and DDX23 at R-loop-containing loci

Co-transcriptional R-loops impact on RNA Pol II transcription dynamics. For instance, they are involved in the promoter-proximal pausing and promote efficient transcription termination by slowing down RNA Pol II, therefore facilitating the timely recruitment of RNA processing factors (Ginno et al., 2012; Skourti-Stathaki et al., 2011 Harlen et al., 2016; Nojima et al., 2015). For this reason, we sought to investigate whether oscillations in transcription reach back to 'signal' the suppression of persistent R-loops.

In order to investigate RNA Pol II dynamics at the site of R-loop formation, we induced R-loops in cells by depleting *THOC1*. In agreement with the view that R-loops slow down transcription, we detected a significant accumulation of RNA Pol II in *APOE* and *ACTB* loci where R-loops had been detected, but not in the *SNRPN* locus, upon *THOC1* knockdown (**Fig.26a**). *THOC1* knockdown did not cause significant changes in the expression levels of *APOE* and *ACTB* (**Fig.26c**), suggesting that the observed RNA Pol II accumulation is not the result of changes in the transcription rate, but rather a localized pausing event. Indeed, the RNA Pol II accumulated at the *APOE* and *ACTB* loci carried a serine 5-phosphorylation (Ser5P) pattern (**Fig.26b**), which is suggestive of paused transcription complexes (Buratowski, 2009). Notably, the stalled RNA Pol II complexes were released upon suppression of R-loops by RNase H1 overexpression (**Fig. 26a,b**).

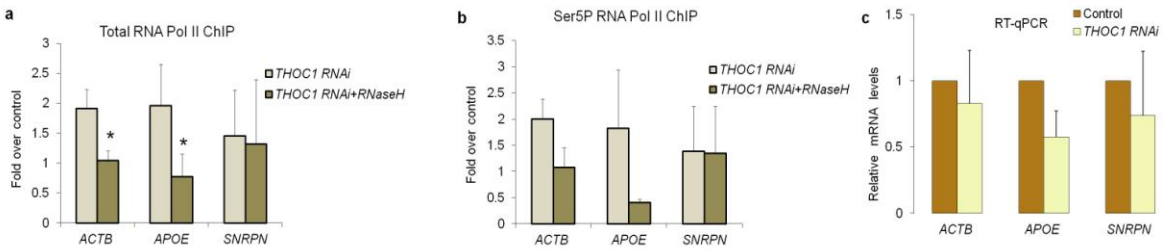


Figure 26. RNA Pol II pauses at R-loop-containing sites. ChIP analysis of RNA Pol II **(a)** and Ser5p RNA Pol II **(b)** in *THOC1* RNAi-depleted cells with (+ RNase H1) or without transient transfection of *RNase H1*. Data was normalized against the ChIP values obtained in control cells (i.e. without *THOC1* knockdown). Means and standard deviations from three independent ChIP experiments are plotted. Statistical significance was determined using two-tailed Student's t-test. *p < 0.05 **p<0.005. **c)** Total mRNA levels of *ACTB, APOE* and *SNRPN* in U-2OS cells treated with control and *THOC1* RNAi. Data are relative to the *U6 snRNA* expression and was normalized against the RT-qPCR values obtained in control cells. Means and standard deviations are from a minimum of three independent experiments.

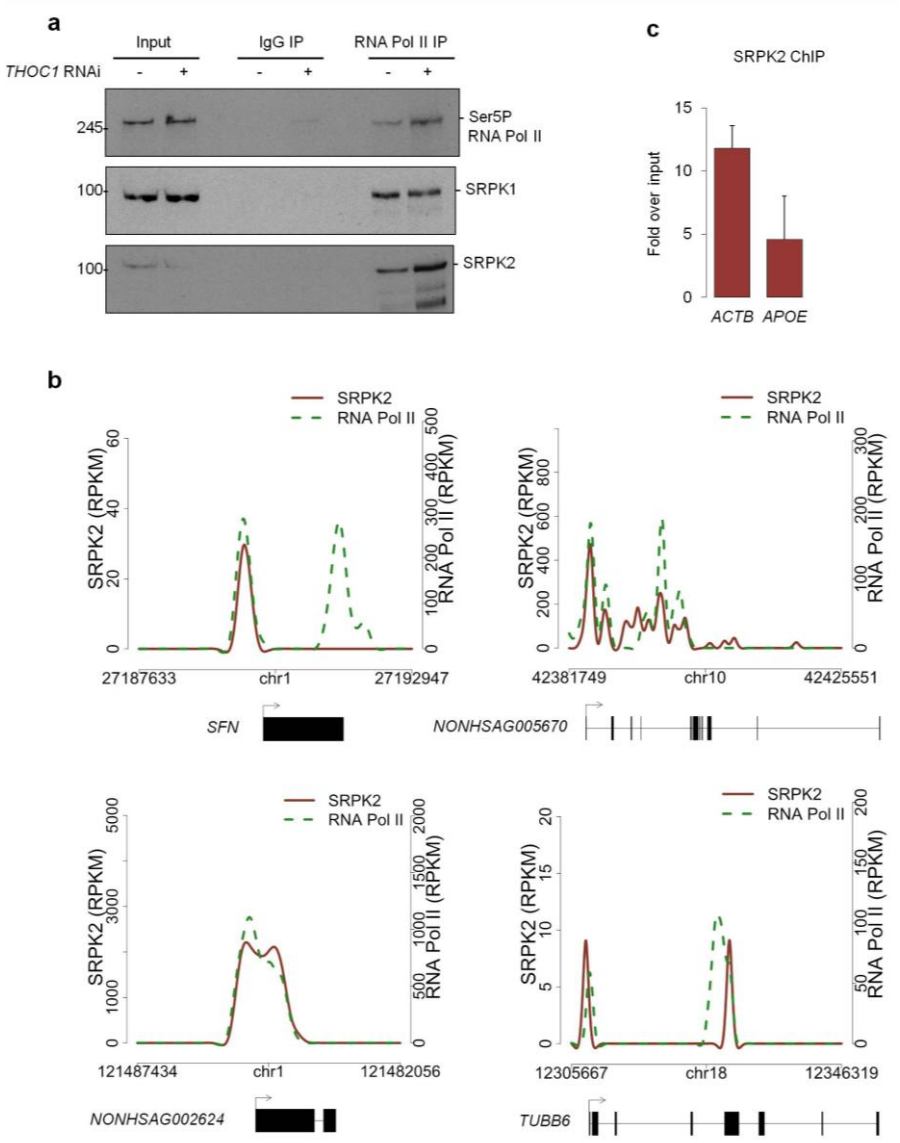


Figure 27. RNA Pol II pausing nucleates SRPK2 at R-loop-containing sites. a) Nuclear immunoprecipitations of RNA Pol II on control and *THOC1* RNAi depleted U-2OS cells. Purified complexes were resolved by SDS-PAGE and blotted with antibodies against Ser5P-RNA Pol II, SRPK1 or SRPK2. The Input lane represents total cell lysates and IgG IP denotes the negative control immunoprecipitation. Molecular weight markers (KDa) are shown on the left. Data are representative of three independent experiments. **b)** Individual profiles with 20 bp windows of RNA Pol II and SRPK2 distribution (RPKMs) along four different loci. **c)** ChIP analysis of SRPK2 in U-2OS cells. Data were normalized against the input values. Means and standard deviations are from a minimum of three independent ChIP experiments.

Given the intimate link with R-loops, we reasoned that stalled RNA Pol II complexes would constitute ideal molecular sensors to 'signal' the location of intragenic R-loops and initiate a molecular pathway towards their suppression. According to our model, paused RNA Pol II would nucleate SRPK2-dependent phosphorylation of DDX23. In agreement with our view, co-immunoprecipitation experiments revealed that the accumulation of R-loops upon *THOC1* knockdown increased Ser5P RNA Pol II levels and promoted the interaction of RNA Pol II with SRPK2, but not with SRPK1 (**Fig.27a**). Moreover, the analysis of ChIP-seq data revealed that SRPK2 is enriched at sites of RNA Pol II pausing (**Fig.27b**). Finally, we observed that SRPK2 coincides with DDX23 at *APOE* and *ACTB* loci (**Fig.27c**). Altogether, these data strengthen our **proposed model whereby changes in transcription dynamics initiate a molecular pathway leading to SRPK2-dependent DDX23 phosphorylation and suppression of R-loops**.

Imagination meets memory in dark.

Annie Dillard

Discussion primes ideas, argument primes conflicts.

Anonymous

Every great magic trick consists of three parts or acts. The first part is called "The Pledge". The magician shows you something ordinary: a deck of cards, a bird or a man. He shows you this object. Perhaps he asks you to inspect it to see if it is indeed real, unaltered, normal. But of course... it probably isn't. The second act is called "The Turn". The magician takes the ordinary something and makes it do something extraordinary. Now you're looking for the secret... but you won't find it, because of course you're not really looking. You don't really want to know. You want to be fooled. But you wouldn't clap yet. Because making something disappear isn't enough; you have to bring it back. That's why every magic trick has a third act, the hardest part, the part we call "The Prestige".

Opening scene from 'The Prestige'- Chris Nolan

Atticus told me to delete the adjectives and I'd have the facts.

Harper Lee

Facts are many, but the truth is one.

Tagore

Chapter 5. Discussion

Highlights

- RNA Pol II 'senses and signals' R-loops and nucleates SRPK2 as the 'signal transducer' that will phosphorylate DDX23 − 'the effector' − to maintain R-loop homeostasis.
- The genomic instability resulting from *DDX23* deficiency may explain the frequent high-level losses in 12q13 − the genomic housing of *DDX23* – observed in ACC of the salivary glands.

5.1 SRPK2 and DDX23 – new players that prevent RNA-mediated genome instability

SRPK1 and SRPK2 can be categorized as 'stress response' kinases whose localization, activity and functional outcome changes based on different stress signals (Zhong et al., 2009). Available data strongly suggest their role in changing alternative splicing decisions by phosphorylating different SR proteins (Twyffels et al., 2011). Although there is an overlap between the substrates of SRPK1 and SRPK2, they seem to have different functional outcomes (Varjosalo et al., 2013). Indeed it is surprising that only SRPK2 plays a crucial role in genome instability, as described in this thesis. This intriguing result is followed by the fact that SRSF1 phosphomimic was not able to rescue the DNA damage phenotype suggesting the functioning of a new SRPK-SR protein phosphorylation switch.

Additional post-translational modifications could even add to the complexity of SRPK2 function. For instance, histone acetyltransferase Tip60 acetylates the RNA recognition motif of SR splicing factor 2 (SRSF2) and interferes with SRSF2 phosphorylation by inhibiting the nuclear translocation of SRPK2 (Edmond et al., 2011). This acetylation/phosphorylation balance controls the activity of SRSF2. Tip60 was recently shown to bind to R-loop-containing loci in mouse embryonic stem cells (Chen et al., 2015). In agreement, depletion of R-loops decreased the binding of Tip60 to chromatin and increased the polycomb repressive complex 2 occupancy genome wide impairing stem cell differentiation (Chen et al., 2015). Interestingly, the gene encoding Tip60 is a tumor suppressor and is frequently mutated in head and neck squamous cell carcinoma, breast carcinoma and lymphomas (Gorrini et al., 2007). *Whether Tip60 and SRPK2 orchestrate*

the acetylation/phosphorylation switch in addtional SR proteins (namely DDX23) to regulate R-loop levels in order to prevent cancer development is yet to be investigated.

We also observed different patterns of intra-nuclear distribution of SRPK1 and SRPK2. Although both kinases exhibit typical speckles-like pattern commonly exhibited by splicing proteins, SRPK2 has a stronger nucleolar signal. Strikingly, R-loops tend to form or to persist longer in nucleoli as a result of ribosomal DNA transcription by RNA polymerase I (El Hage et al., 2010). Additionally, we also observed a robust increase in nucleolar R-loops in cells depleted of *SRPK2*. This opens up a new question on *what is the role of SRPK2 in RNA Pol I transcription, if any?* However, considering the fact that triptolide (specific inhibitor RNA Pol II) was able to reduce DNA damage levels significantly in *SRPK2* RNAi treated cells we can speculate that SRPK2 has a more prominent role in preventing RNA Pol II-derived-R-loops. Interestingly, RNA Pol II transcripts originating from intronic *Alu* elements (aluRNAs) are enriched in the nucleolus and associate with nucleolar proteins. Inhibiting RNA Pol II transcription disrupts the association of *alu*RNAs with nucleoli and severely challenges nucleolar organization (Caudron-Herger et al., 2015). *alu*RNAs escape degradation after pre-mRNA splicing and processing and remain stable as small RNPs (Jády et al., 2012). *Therefore, it will be interesting to see if non-coding RNA species (in this case aluRNAs) generated by RNA Pol II form R-loops in nucleolar regions and SRPK2 plays any role in regulating such trans-R-loops.*

Cells have evolved molecular mechanisms that can suppress R-loops, such as the novel SRPK2-DDX23 signaling axis described in this thesis. DDX23 is part of a family of R-loop suppressors that includes other helicases, such as SETX, AQR, DHX9 and PIF1. However, *how exactly these helicases work at the molecular level is still not clearly known. How do DDX23 and other helicases distinguish dsDNA, dsRNA and RNA-DNA hybrids?* We did observe that the helicase activity of DDX23 is essential for its R-loop suppressing function. Strikingly, the purified yeast orthologue of DDX23, Prp28, has no detectable ATPase or helicase activity *in vitro* (Muhlmann et al., 2014), even though both activities are required for splicing *in vivo* (Chang et al., 1997).This suggests the need of accessory factors that work together with DDX23 activating it. An example of that is herein illustrated by the role of the SRPK2-dependent phosphorylation of DDX23, which

is required for its integration in the tri-snRNP complex and proficient splicing mechanism (Mathew et al., 2008).

5.2 Potential link between DDX23 and adenoid cystic carcinoma

The finding that *DDX23*-depleted cells exhibit a marked genome instability phenotype strongly suggests a role in cancer. In agreement with this view we detected *DDX23* homozygous deletions in 10 out of 60 (17%) adenoid cystic carcinoma (ACC) samples. ACC is a rare form of malignant neoplasm that arises within secretory glands. The development of the disease is fairly slow yet it follows a steady relentless course. Perineural invasion and high rates of distant metastasis to organs like lungs, liver and bone are characteristic features (Fordice et al., 1999; Spiers et al., 1996). Like for other cancers, the ACC genome exhibit high levels of genome instability. Except for the *MYB-NFIB* gene fusion there are no ACC characteristic genetic abnormalities known so far.

A recent cohort of 60 ACC samples (55 exome sequencing, 5 genome sequencing) with matched normal samples revealed that mutations are enriched in pathways involved in chromatin remodeling, DNA damage response, protein kinase A and phosphatidyl inositol 3-kinase signaling (Ho et al., 2013). Moreover recurrent and high-level losses were also identified in 6q24, 12q13 and 14q loci. Remarkably, 12q13 is the genomic location of DDX23, anticipating a link between ACC development and the R-loop-dependent genome instability.

5.3 Role of R-loops in defining gene boundaries

R-loops are widely distributed along the genome, with a tendency to localize at promoter and terminator regions of RNA Pol II transcribed genes (Skourti-Stathaki and Proudfoot, 2014). R-loops at 3’ end of genes act to slowdown RNA Pol II favoring the recruitment of termination factors and an efficient transcription termination (Skourti-Stathaki et al., 2011). Perturbation in such a mechanism might lead to termination defects resulting in RNA Pol II transcription past the canonical termination site, a feature termed ‘read-through’.

Recent work from our lab strongly suggests that, dysfunction of SET Domain Containing 2 (SETD2) leads to transcriptional read-through up to 4 Kb downstream termination sites (Grosso et al., 2015). SETD2 seeds three methyl groups on lysine 36 of histone 3 (H3K36me3) in RNA Pol II transcribed genes, a mark that is required for the recruitment of the histone chaperone complex FACT during transcription (Carvalho et al., 2013). Notably, deficiency of FACT during transcription leads to R-loop-driven genome instability (Herrera-Moyano et al., 2014). Moreover, our group and others showed that loss of SETD2 increases DNA damage within transcriptional units (Aymard et al., 2014; Pfister et al., 2014). Interestingly, γH2AX and H3K36me3 co-exist in the same nucleosomes and help in the recruitment of DNA repair factors to DNA lesions located in transcriptional units (Carvalho et al., 2014). Altogether, these data suggest a model whereby H3K36me3 deposited during transcription acts to prevent genome instability through a mechanism that may involve FACT-dependent suppression of R-loops.

5.4 Stalled RNA Pol II − a sensor for R-loops

Given that unscheduled R-loop formation compromises the genome stability, dedicated cellular mechanisms must be on call to react promptly in such cases (Sollier and Cimprich, 2015). The work presented in this thesis, describes a mechanism involving RNA Pol II as the *R-loop sensor*, SRPK2 as the *signal transducer* and DDX23 as the *molecular effector* to ward off against deleterious intragenic R-loops. We detect DDX23 accumulation at chromatin loci where R-loops form, which are also co-occupied by Ser5P RNA Pol II and SRPK2. Given the association between Ser5P RNA Pol II and transcription pausing (Alexander et al., 2010; Nojima et al., 2015), we suggest that the concomitant accumulation of SRPK2 and DDX23 onto chromatin follows R-loop-dependent RNA Pol II pausing. While we cannot exclude that SRPK2 accumulates on transcription pause sites, at least partially due to their general role in phosphorylating mRNA processing factors (Ji et al., 2013; Lin et al., 2008; Pandit et al., 2013), its increased association with RNA Pol II observed in *THOC1*-depleted cells suggests a specific recruitment to chromatin in response to R-loops accumulation. We reason that the accumulation of Ser5P RNA Pol II – predominant near the transcription start site (Heidemann et al., 2013) – within the gene body challenges the expected phosphorylation pattern of elongating polymerases and

signals the recruitment of both SRPK2 and DDX23. A similar mechanism linking Ser5P RNA Pol II and the recruitment of RNA-processing factors may explain the splicing-dependent transcription pausing (Harlen et al., 2016; Nojima et al., 2015). "*What are the differences, if any, between an R-loop-dependent and a splicing-dependent RNA Pol II pausing?*" and "*how does transcription pausing signal the nucleation of distinct molecular machineries?*" are important questions that emerge from our model. Importantly, we should also consider the hypothesis that R-loops may form in response to the splicing-dependent RNA Pol II pausing, as was recently shown in the case of transcription-blocking DNA lesions (Tresini et al., 2015). In this scenario, the SRPK2-DDX23 axis described here would be a good candidate to suppress such opportunistic R-loops. Further studies aimed at clarifying these open questions will certainly provide valuable insights into the mechanistic coupling between transcription, pre-mRNA processing and genomic instability.

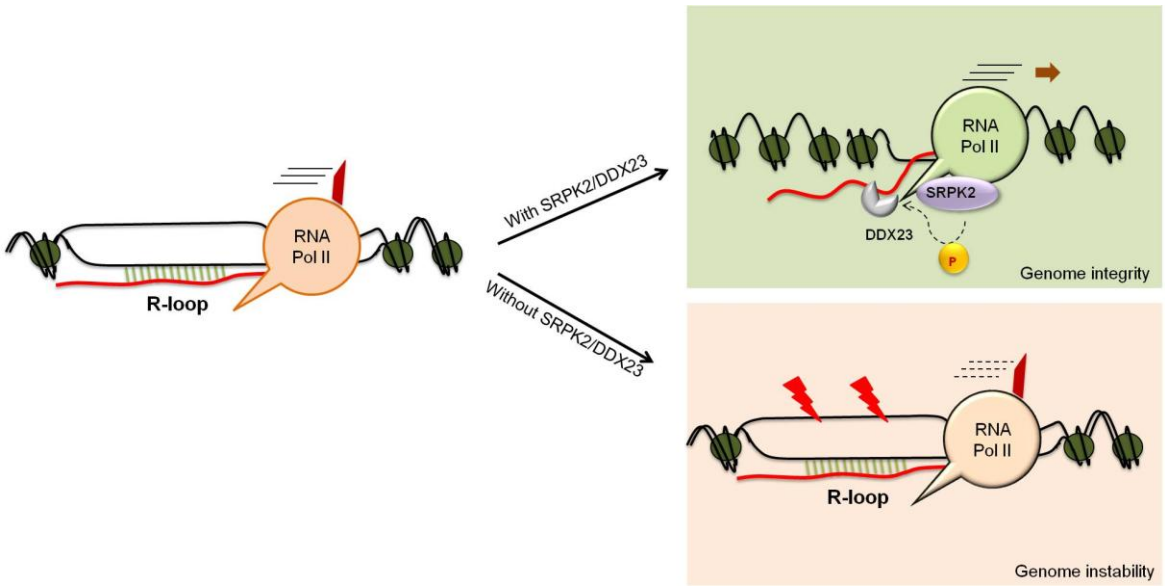


Figure 28. Proposed model for the role of SRPK2-DDX23 in the regulation of R-loops. RNA Pol II pausing (orange) recruits SRPK2 (purple) that phosphorylates DDX23 (gray) to prevent aberrant accumulation of R-loops and maintain the genome integrity. In the absence of either SRPK2 or DDX23, the presence of R-loops favors DNA damaging events (e.g. through collisions with the replication machinery) that create genomic instability.

5.5 Concluding remarks

The work presented in this thesis reveals that pausing of RNA Pol II initiates a signaling cascade whereby SRPK2 phosphorylates DDX23 culminating in the suppression of R-loops. We show that in the absence of either SRPK2 or DDX23, accumulation of R-loops leads to massive genomic instability. Importantly, we detected homozygous deletions of the entire *DDX23* locus in ACC. Our results unravel the molecular details of a novel link between transcription dynamics and RNA-mediated genomic instability that may play important roles in cancer development. Altogether, our data allow us to propose a model whereby R-loops trigger the RNA Pol II pausing, which in turn nucleates SRPK2-dependent DDX23 phosphorylation (**Fig.28**). This model implies that DDX23 activation follows the accumulation of R-loops, placing this helicase as a suppressor rather than a preventer of R-loops. This separates it from the role of several other RNA processing factors, such as SRSF1 or THOC1 that prevent the formation of co-transcriptional RNA-DNA hybrids. Instead, DDX23 would fall within the family of R-loop suppressors, on which the helicase SETX occupies a prominent position. Notably, while SETX resolves RNA-DNA hybrids to promote transcription termination (Skourti-Stathaki et al., 2011), we suggest that DDX23 acts to release RNA Pol II from R-loop-mediated pausing throughout the gene body during transcription elongation.

To fight one might need to know technique, but to hunt one need to use instincts.

Anonymous

What one fool can do another can do better!

Feynman

At some point, everything's gonna go south on you... everything's going to go south and you're going to say, this is it. This is how I end. Now you can either accept that, or you can get to work. That's all it is. You just begin. You do the math. You solve one problem... and you solve the next one... and then the next. And if you solve enough problems, you get to come home….

In the face of overwhelming odds, I'm left with only one option, I'm gonna have to science the shit out of this.

Andy Weir

Appendix A.
Materials & Methods

Cell culture

U-2OS, HeLa and MCF7 cells were grown as monolayers in Dulbecco's modified Eagle medium-DMEM (Invitrogen, Carlsbad, CA), supplemented with 10% (v/v) FBS, 1% (v/v) non-essential amino acids, 1% (v/v) L-glutamine and 100 U/ml penicillin-streptomycin. MCF10A cells were cultured in DMEM/F12 (Invitrogen, Carlsbad, CA) supplemented with 5% (v/v) horse serum, 100µg/ml EGF, 1mg/ml Hydrocortisone, 1mg/ml cholera toxin, 10mg/ml insulin and 100U/ml penicillin-streptomycin (Debnath et al., 2003). All cells were maintained at 37°C in a humidified atmosphere with 5% CO_2.

RNA interference

RNAi was achieved using synthetic siRNA duplexes. siRNAs targeting the firefly luciferase (GL2) were used as controls (Eurogentec, Seraing, Belgium). Cells were reverse transfected with 10 µM siRNAs using OptiMEM (Invitrogen) and Lipofectamine RNAiMAX (Invitrogen), according to the manufacturer's instructions. The sequence of the siRNAs is shown in **Material and Methods Table 1.**

Plasmid construction and transfections

DDX23 expression plasmids were obtained upon modification of *pET21a(+)-DDX23* (kindly given by Ralf Ficner) as follows: *pEGFP-DDX23WT* was constructed by ligating XhoI+XbaI digested *pET21a(+)-DDX23* insert fragment into NheI+XhoI digested *pEGFP-N1* vector. In order to make *pEGFP-DDX23PhosM*, a 715b fragment with nine mutations (S23D, S49D, S57D, S59D, S61D, S63D, S65D, S106D, S118D) and NheI and Van91I sites was synthetically constructed using GeneArt™ technologies and cloned into *pEGFP-DDX23WT*. Similarly for *pEGFP-DDX23PhosIA*, a 715b fragment with nine mutations (S23A, S49A, S57A, S59A, S61A, S63A, S65A, S106A, S118A) and NheI and Van91I sites was synthetically constructed using GeneArt™ technologies and cloned into *pEGFP-DDX23WT*. *pEGFP-DDX23PhosM* was later digested using XcmI and self-ligated to get helicase truncated mutant *pEGFP-DDX23PhosMΔHel*. Plasmids were transfected with supplier's protocol from Lipofectamine 3000 (Invitrogen). *RNaseH1-mCherry* expression was induced with 2.5 µg/ml doxycycline (Sigma, St Louis, MI). The list of all plasmids used in this study is shown in **Material and Methods Table 1.**

Western blot and sub-cellular fractionation

Whole cell protein extracts were prepared by cell lysis with SDS-PAGE buffer (80 mM Tris-HCL pH 6.8, 16% glycerol, 4.5% SDS, 450 mM DTT, 0.01% bromophenol blue) with 200 U/ml benzonase (Sigma) and 50 μM $MgCl_2$ and boiling for 5 min. For cellular fractionation, U-2OS were swelled and fractioned to separate cytosolic fraction and nuclei (De Almeida et al., 2010). Whole nuclei were then lysed in RIPA buffer (50 mM Tris pH=8.0, 150 mM NaCl, 1% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS) and centrifuged for 10 min at 15000 rpm to isolate all the nucleoplasmic proteins. The chromatin fraction was digested with 20U DNase I. Proteins from the different fractions were precipitated using standard trichloro acetic acid-acetone method (TCA-acetone method). Equal amounts of protein extracts were resolved by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a nitrocellulose membrane. Details of antibodies used are mentioned in **Material and Methods Table 1.**

Isolation of chromatin with RNase A digestion

Nuclei from HeLa cells were isolated as previously described (De Almeida et al., 2010). Briefly, HeLa cells were swelled and fractioned to separate cytosolic fraction and nuclei. Whole nuclei were re-suspended in 1X RNase reaction buffer (20 mM Tris-HCl (pH=8.4), 50 mM KCl, 5 mM $MgCl_2$, 20mM DTT) and divided into two halves. One half was digested with 1.5U RNase A (Roche) for 90min at 37^0C and another with buffer alone. RNase A digested and undigested nuclei were re-suspended (5% Glycerol, 20mM Tris (pH=7.9), 75mM NaCl, 0.5mM EDTA, 0.85mM DTT, 0.125mM PMSF) and lysed (20mM HEPES (pH=7.6), 300mM NaCl, 0.2mM EDTA, 1mM DTT, 7.5mM $MgCl_2$, 1M Urea, 1% NP40) on ice for 10min. The samples were centrifuged for 10 min at 15000 rpm to isolate chromatin. The step is repeated to isolate pure chromatin. The chromatin fraction was digested with 20U DNase I. Proteins from different fractions were precipitated using the standard TCA-acetone method. Equal amounts of protein extracts were resolved by SDS-PAGE and transferred to a nitrocellulose membrane. Details of antibodies used are mentioned in **Material and Methods Table 1.**

Immunofluorescence

U-2OS cells grown on coverslips were fixed with 3.7% paraformaldehyde for 10 min at room temperature. The cells were then permeabilized with 0.5% Triton X-100/PBS for 10 min and

incubated with primary antibodies against γH2AX, cyclin A, SRPK1 and SRPK2 (Carvalho et al., 2014). R-loops were detected with the S9.6 antibody following cell fixation and permeabilization with 100% ice-cold methanol and acetone for 10 min and 1 min on ice. Incubation with primary antibodies was followed by incubation with fluorochrome-conjugated secondary antibodies. All the washing steps were done with PBS containing 0.05% (v/v) Tween 20. The samples were mounted in Vectashield (Vector Laboratories, Burlingame, CA) with 4'-6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich) to stain the DNA. Images were taken using a Zeiss LSM 710 (Carl Zeiss, Oberkochen, Germany) confocal microscope with a 63x/1.4 oil immersion and quantified with Image J. List of primary and secondary antibodies used are mentioned in **Material and Methods Table 1.**

Analysis of metaphase spreads

Metaphases were prepared from HeLa cells as described (Gallego-Paez et al., 2014) with the following modifications. Briefly, five days after RNAi cells were synchronized with 1mM thymidine (Sigma) for 20 hours and released from thymidine block for 6 hours. Cells were then treated with 1μg/ml nocodazole (Sigma) for 2 hours and mitotic cells were collected by mitotic shake. Cells were then centrifuged and re-suspended in pre-warmed 75 mM KCl for 5 min and fixed with freshly prepared ice-cold Carnoy's solution (3:1 methanol:acetic acid solution). Fixed chromosomes were washed and re-suspended in Carnoy's solution. Chromosomes were then dropped on clean glass side, dried and stained with Giemsa (Millipore). Excess Giemsa stain was washed with distilled water and glass slides were dried. Dried chromosome spreads were mounted using Fluoromount-G medium (Southern Biotech, Birmingham, AL, USA). Randomly selected metaphase spreads were imaged using 100X oil immersion (Leica DM2500 microscope). Chromosomal aberrations were quantified on three independent experiments. At least 20 metaphases per condition were analyzed on each independent experiment

Cell cycle analysis

The nuclear DNA content was estimated by flow cytometry analysis of cells stained with propidium iodide (PI) as described earlier (Carvalho et al., 2014). Briefly, U2-OS cells transfected with *RNase H1-GFP* or treated with triptolide were collected by trypsinization and fixed by drop-wise addition of ice-cold 50% ethanol with gentle vortexing followed by RNA digestion and addition of PI (100 μg/ml). Acquisition of cells transfected with *RNase H1-GFP* was performed on

a FACS Calibur (BD Biosciences) and acquisition of cells treated with triptolide was performed on a BD Accuri C6 (BD Biosciences). Data was processed with FlowJo (TreeStar).

Co-Immunoprecipitation

Co-immunoprecipitations of RNA Pol II with SRPKs were performed on nuclei extracts as described (Carvalho et al., 2014). Briefly, isolated nuclei were lysed (20 mM Hepes pH 7.0, 10 mM KCl, 0.1% (v/v) NP40, 6 mM $MgCl_2$, 20% (v/v) glycerol, protease-(Roche) and phosphatase inhibitors (Roche) for 10mins on ice. Then nuclei were sonicated with a single pulse of 15 s at 20% intensity using a Soniprep150 and digested with 20U DNase I (Roche) at 4°C for 60 min before pre-clearing with Protein G Dynabeads (Life Technologies) at 4°C for 30 min. 1µg of DNase I digested DNA was run on 1% agarose gel to check if DNA is fragmented in the given conditions (**Extended Data Fig.2d**). Samples were diluted in IP buffer (20 mM Hepes pH 7.0, 10 mM KCl, 1.5 mM $MgCl_2$, 0.2%(v/v) Tween20, 10% (v/v) glycerol, 1 mM DTT) and incubated with respective antibodies overnight at 4°C. The protein complexes were pulled down using Protein G Dynabeads for 2h at 4°C, washed one time in wash buffer (20 mM Hepes pH 7.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2% (v/v) Tween20, 10% (v/v) glycerol, 1 mM DTT) with increasing KCl concentrations of 50 mM, 100mM and 150 mM. The final wash was made in IP buffer. Protein samples were then eluted in 2X Laemmli buffer and resolved by SDS-PAGE before immunobloting. 1/10th of the total cell lysate was used as input samples. Overexpressed DDX23-WT or DDX23-PhosM-ΔHel proteins were immunoprecipitated from U2-OS whole-cell lysates with a T7 antibody (Millipore, AB3790). Before sonication, cells were cross-linked for 5 min with 1% formaldehyde followed by quenching with 1M glycine. Pulled down complexes were resolved by western blot. List of antibodies used are mentioned in **Material and Methods Table 1**.

DRIP

DNA-RNA hybrids were immunoprecipitated as described earlier (Ginno et al., 2012) with following modifications. Briefly, U2-OS cells transfected with control, *DDX23* and *THOC1* RNAi were collected after 72hr and lysed in lysis buffer (100 mM NaCl, 10 mM Tris pH 8.0, 25 mM EDTA pH 8.0, 0.5% SDS, 50µg/ml Proteinase K) for 5hr at 45^0C. Nucleic acids were extracted using standard phenol-chloroform extraction protocol and re-suspended in DNase/RNase-free water. Nucleic acids were then fragmented using restriction enzyme cocktail (20U each of EcoRI, BamHI, HindIII, BsrgI and XhoI). Half of the sample was digested with 40U RNase H (NZYtech)

to serve as a negative control, for about 36-48hr at 37^0C. Digested nucleic acids were cleaned with standard phenol-chloroform extraction method and res-suspended in DNase/RNase-free water. DNA-RNA hybrids were immunoprecipitated from the total nucleic acids using 6.5µg of S9.6 antibody in binding buffer (10 mM $NaPO_4$ pH 7.0, 140 mM NaCl, 0.05% Triton X-100) overnight at 4^0C. 50µl Protein A/G magnetic beads (Pierce™, Thermo scientific) were used to pull down the immunoprecipitates at 4^0C for 2-3hrs. Pulled-down complexes were washed twice with binding buffer, once using TE buffer each for 15min at RT, followed by elution (50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS, 2.5µg proteinase k) for 30 min at 55^0C. Nucleic acids were extracted using phenol-chloroform method. Restriction enzyme digested nucleic acid material was used as input. The relative occupancy of the immunoprecipitated DNA-RNA hybrids at each locus was estimated by RT-qPCR as follows: $2^{(CtIn- CtIP)}$, where ^{CtIn} and ^{CtIP} are mean threshold cycles of RT-qPCR done in duplicate on samples from input and immunoprecipitations, respectively. The sequence of the primers is shown **Material and Methods Table 1**.

RNA isolation and quantitative RT-PCR

RNA isolation and cDNA preparation was made as described earlier (Grosso et al., 2015). Briefly, total RNA was isolated from U2-OS cells transfected with control or *THOC1* siRNAs for 48h using TRIzol (Invitrogen). cDNA was made using Superscript II Reverse Transcriptase (Invitrogen). RT-qPCR was performed in the ViiA Real Time PCR (Applied Biosystems, CA, USA), using SYBR Green PCR master mix (Bio-Rad). The relative RNA expression was estimated as follows: $2^{(Ct\ reference - Ct\ sample)}$, where Ct reference and Ct sample are mean threshold cycles of RT-qPCR done in duplicate of the U6 snRNA (reference) and the gene of interest (sample). All primer sequences are presented in **Material and Methods Table 1**.

ChIP-qPCR and ChIP-seq

Chromatin immunoprecipitation (ChIP) was performed in U-2OS cells as described (de Almeida et al., 2011; Carvalho et al., 2013). Antibodies used for ChIP were mentioned in **Material and Methods Table 1**. The relative occupancy of the immunoprecipitated protein at each DNA locus was estimated by RT-qPCR as follows: $2^{(Ct\ Input - Ct\ IP)}$, where Ct Input and Ct IP are mean threshold cycles of RT-qPCR done in duplicate on DNA samples from input and specific immunoprecipitations, respectively. The sequence of the primers is shown in **Material and Methods Table 1**. ChIP-seq was performed as described (de Almeida et al., 2011). Briefly, ~80%

confluent non-transformed human mammary epithelial cells (MCF10A), were incubated with 1% (v/v) formaldehyde for 10 min to perform cross-linking and quenched with 250 mM glycine for 5 min. Cells were lysed and sonicated to shear chromatin to get 100-300bp fragments. 1µg of input DNA was run on 1% agarose gel to check if most of the fragmented DNA lie within the range of 100-300bp. Pre-cleared chromatin was incubated separately with antibodies against SRPK1 (ab90527, Abcam) or SRPK2 (A302-467A, Bethyl Laboratories) overnight at 4^0C. ~$1/10^{th}$ of the sample was taken aside as inputs. Immunoprecipitated and input DNA were subsequently purified and quantified. For biological replicates, all steps were repeated using independent samples. Libraries were prepared with DNA from two biological replicates according to the Illumina protocol and sequenced at the GeneCore genomics service center at the EMBL.

Bioinformatics Analyses

RNA-seq, GRO-seq and RNA Pol II ChIP-seq data for MCF10A were obtained from the Sequence Read Archive (RNA-seq: SRX685939; GRO-seq: ERR034678, ERR039508, ERR039509; RNA Pol II ChIP-seq: SRX143301). The quality of HTS data was assessed with FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc). ChIP-seq reads were aligned to the reference human genome (GRCh37/hg19 assembly) with Bowtie (Langmead et al., 2009) and filtering for uniquely aligned reads. Enriched regions were identified for each individual replicate using MACS (Zhang et al., 2008), with a false-discovery rate of 0.05. Only genomic regions consistently identified between biological replicates were considered for downstream analyses. Finally, ChIP-seq enriched regions were assigned to annotated genes, including a 4 kb region upstream the transcription start site and downstream the transcription termination site (**Extended Data Table 1**). Gene annotations were obtained from Ensembl (GRCh37.75 version (Flicek et al., 2014)) and merged into a single transcript model per gene using BedTools (Quinlan and Hall, 2010). For individual and metagene profiles, uniquely mapped reads were extended in the 3' direction to reach 150 nucleotides with the Pyicos (Althammer et al., 2011). For the metagene profile, genes were aligned at the first and last nucleotides of the annotated transcripts and read counts were scaled as follows: the 5' end (10 kb upstream of the transcription start site) and the 3' end (10 Kb downstream of the transcription termination site) were unscaled and averaged in a 100 bp window, and the remainder of the gene was represented by 200 values from cubic spline interpolation so that all genes seem to have the same length. Individual profiles were produced using a 20 bp window. All profiles were plotted on a normalized reads per kb per million mapped reads (RPKMs). Expression levels (TPMs) from RNA-seq and GRO-seq datasets were obtained using Kallisto (Bray et al., 2016). Transcriptionally active genes were defined as those with

expression levels higher than the 25th percentile.

GC features were assessed for genic SRPK1 and SRPK2 binding sites co-oriented with transcription and recovered 2 kb of DNA sequence on each side of the peak summit. GC skew (calculated using the formula (G-C)/(G+C)) and CpG density (measured as the CpG observed versus expected ratio or CpG o/e) were computed for 50 bp sliding windows of 1 bp step size. The random peak dataset was generated using the shuffle Bedtools function, by permutating the SRPK2 peaks (maintaining the number and length) from annotated genes.

Point-mutations in DDX23 were obtained from cBioportal (Gao et al., 2013) their effect was predicted using PredictSNP (Bendl et al., 2016) (**Extended Data Table 2**).

A set of in-house scripts for data processing and graphical visualization were written in bash and in the R environmental language http://www.R-project.org. SAMtools (Li et al., 2009) and BED tools were used for alignment manipulation, filtering steps, file format conversion and comparison of genomic features.

Material and Methods Table 1.

Product	Concentrations/ incubation time	Company/Cat.No.	Notes
Drugs	**Final concentration**		
Triptolide	500nM/70min	Santa Cruz; sc-200122	To inhibit RNA Pol II transcription inhibiton
Nocodazole	1µg/ml	Sigma-Aldrich; M1404	To arrest cells in mitotic phase
Doxycycline	2.5µg/ml for 48hr	Sigma-Aldrich; D9891	To induce RNaseH1-mCherry plasmid
siRNA			
Control siRNA	10µM/48-72hr	Eurogentec	siRNA against fire fly luciferase locus Sense: CGUACGCGGAAUACUUCGA Anti-sense: UCGAAGUAUUCCGCGUACG
***SRPK1* siRNA**	10µM/48-72hr	Santa Cruz; sc-39235	To deplete *SRPK1* transiently
***SRPK2* siRNA**	10µM/48-72hr	Santa Cruz; sc-3923 Origene; SR304591 Qiagen; SI00162925, SI00162932	To deplete *SRPK2* transiently
***DDX23* siRNA**	10µM/48-72hr	Santa Cruz; sc-62200 Origene; SR306242 Qiagen; SI03139514, SI03144743	To deplete *DDX23* transiently
***THOC1* siRNA**	10µM/48-72hr	Santa Cruz; sc-76652	To deplete *THOC1* transiently
***PRP8* siRNA**	10µM/48-72hr	Santa Cruz; sc-38209	To deplete *PRP8* transiently
***PRP6* siRNA**	10µM/48-72hr	Qiagen; SI03101945, SI05021646	To deplete *PRP6* transiently
Plasmids			
pEGFP-RNaseH1	48hr	Provided by Dr. Robert J Crouch	Bacterial RNaseH1 tagged with GFP
pICE-RNaseHI-mCherry	48hr	Provided by Dr. Patrick Calsou	Bacterial RNaseH1 tagged with mCherry inducible by doxycycline.
pCGT-SRSF1WT	48hr	Provided by Dr. Javier Caceres	Human SRSF1 wild type tagged with T7.
pCGT-SRSF1PhosM	48hr	Provided by Dr. Javier Caceres	Human SRSF1 phosphomimic mutant tagged with T7.
pEGFP-DDX23WT	48hr	Self made with advice from Dr. Robert Martin	Human DDX23 wild type, tagged with T7 and GFP.
pEGFP-DDX23PhosM	48hr	Self made with advice from Dr. Robert Martin	Human DDX23 phosphomimic mutant, tagged with T7 and GFP.
pEGFP-DDX23PhosIA	48hr	Self made with advice from Dr. Robert Martin	Human DDX23 phospho inactive mutant, tagged with T7 and GFP.
pEGFP-DDX23PhosMΔHel	48hr	Self made with advice from Dr. Robert Martin	Human DDX23 phosphomimic mutant with helicase truncation, tagged with T7 and GFP.
Antibodies			
S9.6	1µg/25µl (IF), 6.5µg/IP	Kerafast; ENH001	RNA-DNA hybrid antibody used to detect R-loops
γH2AX	1:2500 (WB), 1:200 (IF)	Millipore; 05-636	Serine 139 phosphorylated H2AX, DNA damage marker
SRPK1	1:2000 (WB), 1:200 (IF), 2-4µg/IP	Abcam; ab90527 BD Bioscience; 611072	Serine-Arginine protein kinase 1 specific antibody
SRPK2	1:2000 (WB), 1:200 (IF), 2-4µg/IP	BD Bioscience; 611118 Bethyl Laboratory; A302-467A	Serine-Arginine protein kinase 2 specific antibody
DDX23	1:500 (WB), 2-3µg/IP	Santa Cruz; sc-133504	DEAD box helicase 23 specific antibody
THOC1	1:1000 (WB)	Genetex; GTX70220	THO complex 1 specific antibody
RNA Pol II (N20)	1:200 (WB), 3µg/IP	Santa Cruz; sc-899	Total RNA Pol II antibody
Ser5P-RNA Pol II	1:250 (WB), 50µl/IP	Provided by Prof. Carmo Fonseca's lab	Serine 5 phoshorylated RNA Pol II antibody
Ser2P-RNA Pol II	50µl/IP	Provide by Prof. Carmo Fonseca's lab	Serine 2 phoshorylated RNA Pol II antibody

SRSF1	1:500 (WB)	Provided by Dr. Adrian Krainer's lab	SRSF1 specific antibody
GFP	1:1000	Roche; 1181446000	To detect GFP (green fluorescent protein) tagged proteins (here DDX23)
U1A-30K	1:1000 (WB)	Abcam; ab55751	Splicing factor, here used as a control for RNase A digestion
Cyclin A	1:200 (IF)	Santa Cruz; sc-751	Marker for replicating cells
T7	1:200 (IF) 3µg/IP	Millipore; AB3790	To detect T7 tagged proteins (here SRSF1 and DDX23)
α-tubulin	1:10000 (WB)	Sigma-Aldrich; T5168	Loading control in immunoblots
H3	1:1000 (WB)	Abcam; ab1791	Loading control in immunoblots
Histone H2A	1:1000 (WB)	Abcam; ab18255	Loading control in immunoblots
Histone H2B	1:1000 (WB)	Abcam; ab1790	Loading control in immunoblots
Cy3 anti-mouse	1:200 (IF)	Bethyl; A90-516C3	secondary antibodies used in immunofluorescence
Cy3 anti- rabbit	1:200 (IF)	Bethyl; A120-201C3	secondary antibodies used in immunofluorescence
Dy 488 anti-mouse	1:200 (IF)	Bethyl; A90-244D2	secondary antibodies used in immunofluorescence
Primers			
APOE	10µM	Sequences provided by Bhatia et al., 2014 Made by Life Technologies	*APOE* R-loop primer Fw: CCGGTGAGAAGCGCAGTCGG Rv: CCCAAGCCCGACCCCGAGTA
		Self designed & made by Life Technologies	Primers to check *APOE* mRNA levels Fw: GGCAGAGCGGCCAGCG Rv: CTCCTCCTGCACCTGCTC
ACTB	10µM	Sequences provided by Skourti-Stathaki et al., 2011 Made by Life Technologies	*ACTB* R-loop primers Fw: TTACCCAGAGTGCAGGTGTG Rv: CCCCAATAAGCAGGAACAGA Fw: GGGACTATTTGGGGGTGTCT Rv: TCCCATAGGTGAAGGCAAAG
		Self designed & made by Life Technologies	Primers to check *ACTB* mRNA levels Fw:GGAGGAGCTGGAAGCAGC Rv:CTGGCCGGGACCTGACT
SNRPN	10µM	Sequences provided by Bhatia et al., 2014 Made by Life Technologies	*SNRPN* R-loop negative locus Fw: TGCCAGGAAGCCAAATGAGT Rv: TCCCTCTTGGCAACATCCA
		Self designed & made by Life Technologies	Primers to check *SNRPN* mRNA levels Fw:GTAATTGGGACTCCCATCAAG Rv:GAGAAGTGCCCCACGTGG
PRP6	10µM	Self designed & made by Life Technologies	Primers to check *PRP6* mRNA levels Ex9 Fw GATGTCTGGCTGGAAGCAG Ex9 Rv GCTCACCTTTCCGAAGAACC Ex6 Fw GCAATTTGGAGGTCTTAACACA Ex6 Rv GCCTCATGTCCATCAGAGTG
RBM15	10µM	Self designed & made by Life Technologies	SRPK1 ChIP-qPCR primers Fw CCTCCGTTATACAGGCCTAC Rv AGAGGCAGGCTCACAGCG
MRP63	10µM	Self designed & made by Life Technologies	SRPK1 ChIP-qPCR primers Fw CACAGCCTCCCGCCACTA Rv CTGCCGCGCTGCTCAAG
ECHS1	10µM	Self designed & made by Life Technologies	SRPK1 ChIP-qPCR primers Fw GCGTGCAGGTCGGAGTCAGGA Rv CCGGGCGAGGAGTCCAGAG
NDRG3	10µM	Self designed & made by Life Technologies	SRPK2 ChIP-qPCR primers Fw GAGCTCTGTACCTGGGTGA Rv CCAGGGGACGAACTTGGA
ERO1B	10µM	Self designed & made by Life Technologies	SRPK2 ChIP-qPCR primers Fw AGTGCTGGGATTACGGTGTAA Rv AGCTCCTCAGTGGAAAGTGA
SENP5	10µM	Self designed & made by Life Technologies	SRPK2 ChIP-qPCR primers Fw GGCGGATCACAAGGTCAG Rv TGAATGGAATCGTCATCG ATGA

Intergenic region	10µM	Self designed & made by Life Technologies	Fw:GGAATAACAGATTGGAGGAGG Rv: GCCTCCTGAACTACATCCC
U6snRNA	10µM	Self designed & made by Life Technologies	Fw: GCTTCGGCAGCACATATACTA Rv: AAATATGGAACGCTTCACGA

*WB, Western blot; IF, Immunofluorescence; IP, Immunoprecipitation or chromatin immunoprecipitation.

**This list contains details of reagents more specific for this study. More general reagents used have been mentioned in 'Material and Method' section when describing respective protocol.

We read the world wrong and say that it deceives us.

Tagore

The adventurous student will always study classics…. it requires training such as the athletes underwent the steady intention almost of the whole life to this object. Books must be read as deliberately and reservedly as they were written.

Henry David Thoreau

Appendix B. References

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R. V, Liu, W., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. Science (80-.). *2085*, 1–11.

Aguilera, A. (2002). NEW EMBO MEMBER'S REVIEW 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.

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

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

de Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., et al. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nat. Struct. Mol. Biol. *18*, 977–983.

De Almeida, S.F., García-Sacristán, A., Custódio, N., and Carmo-Fonseca, M. (2010). A link between nuclear RNA surveillance, the human exosome and RNA polymerase II transcriptional termination. Nucleic Acids Res. *38*, 8015–8026.

Althammer, S., González-vallinas, J., Ballaré, C., Beato, M., and Eyras, E. (2011). Pyicos: A versatile toolkit for the analysis of high-throughput sequencing data. Bioinformatics *27*, 3333–3340.

Amin, E.M., Oltean, S., Hua, J., Gammons, M.V.R., Hamdollah-Zadeh, M., Welsh, G.I., Cheung, M.K., Ni, L., Kase, S., Rennel, E.S., et al. (2011). WT1 Mutants Reveal SRPK1 to Be a Downstream Angiogenesis Target by Altering VEGF Splicing. Cancer Cell *20*, 768–780.

Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T., and Azzalin, C.M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. Nat. Commun. *5*, 1–11.

Aymard, F., Bugler, B., Schmidt, C.K., Guillou, E., Caron, P., Briois, S., Iacovoni, J.S., Daburon, V., Miller, K.M., Jackson, S.P., et al. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. Nat. Struct. Mol. Biol. *21*, 366–374.

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. Nat Struct Mol Biol *20*, 1199–1205.

Ballatore, C., Lee, V.M.-Y., and Trojanowski, J.Q. (2007). Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. Nat. Rev. Neurosci. *8*, 663–672.

Baranello, L., Bertozzi, D., Fogli, M.V., Pommier, Y., and Capranico, G. (2009). DNA topoisomerase I inhibition by camptothecin induces escape of RNA polymerase II from

promoter-proximal pause site, antisense transcription and histone acetylation at the human HIF-1?? gene locus. Nucleic Acids Res. *38*, 159–171.

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*.

Bendl, J., Musil, M., Štourač, J., Zendulka, J., Damborský, J., and Brezovský, J. (2016). PredictSNP2: A Unified Platform for Accurately Evaluating SNP Effects by Exploiting the Different Characteristics of Variants in Distinct Genomic Regions. PLoS Comput. Biol. *12*.

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

Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J.M., and Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. Nature *529*, 551–554.

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

Berti, M., and Vindigni, A. (2016). Replication stress: getting back on track. Nat. Struct. Mol. Biol. *23*, 103–109.

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

Boesler, C., Rigo, N., Anokhina, M.M., Tauchert, M.J., Agafonov, D.E., Kastner, B., Urlaub, H., Ficner, R., Will, C.L., and Lührmann, R. (2016). A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. Nat. Commun. *7*, 11997.

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.

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

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

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Brown, T.A., Tkachuk, A.N., and Clayton, D.A. (2008). Native R-loops persist throughout

the mouse mitochondrial DNA genome. J. Biol. Chem. *283*, 36743–36751.

Buratowski, S. (2009). Progression through the RNA Polymerase II CTD Cycle. Mol. Cell *36*, 541–546.

Calo, E., Flynn, R. a, Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2014). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature.

Carvalho, S., Raposo, A.C., Martins, F.B., Grosso, A.R., Sridhara, S.C., Rino, J., Carmo-Fonseca, M., and De Almeida, S.F. (2013). Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. Nucleic Acids Res. *41*, 2881–2893.

Carvalho, S., Vítor, A.C., Sridhara, S.C., Filipa, B.M., Ana, C.R., Desterro, J.M.P., Ferreira, J., and de Almeida, S.F. (2014). SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. Elife *2014*, 1–19.

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

Castellano-Pozo, M., Santos-Pereira, J., Rondón, A., 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.

Caudron-Herger, M., Pankert, T., Seiler, J., Németh, A., Voit, R., Grummt, I., and Rippe, K. (2015). Alu element-containing RNAs maintain nucleolar structure and function. EMBO J. *34*, 2758–2774.

Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A.R., and Cáceres, J.F. (2002). Nuclear export and retention signals in the RS domain of SR proteins. Mol. Cell. Biol. *22*, 6871–6882.

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.

Chae, Y.K., Chung, S.Y., Davis, A.A., Carneiro, B.A., Chandra, S., Kaplan, J., Kalyan, A., and Giles, F.J. (2015). Adenoid cystic carcinoma: current therapy and potential therapeutic advances based on genomic profiling. Oncotarget *6*.

Chakraborty, P., and Grosse, F. (2011). Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. DNA Repair (Amst). *10*, 654–665.

Champoux, J.J. (2001). DNA topoisomerases: structure, function, and mechanism. Annu. Rev. Biochem. *70*, 369–413.

Chan, C.B., and Ye, K. (2013). Serine-arginine protein kinases: New players in neurodegenerative diseases? Rev. Neurosci. *24*, 401–413.

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*.

Chang, T.H., Latus, L.J., Liu, Z., and Abbott, J.M. (1997). Genetic interactions of conserved regions in the DEAD-box protein Prp28p. Nucleic Acids Res. *25*, 5033–5040.

Chédin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends Genet. *xx*, 1–11.

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

Cheung, A.C.M., and Cramer, P. (2012). A movie of RNA polymerase II transcription. Cell *149*, 1431–1437.

Ciccia, A., and Elledge, S.J. (2010). The DNA Damage Response: Making It Safe to Play with Knives. Mol. Cell *40*, 179–204.

Cleaver, J.E., Laposa, R.R., and Limoli, C.L. (2003). DNA replication in the face of (In)surmountable odds. Cell Cycle *2*, 310–315.

Cloutier, S.C., Wang, S., Ma, W.K., Al Husini, N., Dhoondia, Z., Ansari, A., Pascuzzi, P.E., and Tran, E.J. (2016). Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation. Mol. Cell 1–12.

Colak, D., Zaninovic, N., Cohen, M.S., Rosenwaks, Z., Yang, W.Y., Gerhardt, J., Disney, M.D., and Jaffrey, S.R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science (80-.). *343*, 1002–1005.

Costantino, L., and Koshland, D. (2015). The Yin and Yang of R-loop biology. Curr. Opin. Cell Biol. *34*, 39–45.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Debnath, J., Muthuswamy, S.K., and Brugge, J.S. (2003). Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods *30*, 256–268.

Dietlein, F., Thelen, L., and Reinhardt, H.C. (2014). Cancer-specific defects in DNA repair pathways as targets for personalized therapeutic approaches. Trends Genet. *30*, 326–339.

Domínguez-Sánchez, M.S., Barroso, S., Gómez-González, B., Luna, R., and Aguilera, A. (2011). Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. PLoS Genet. *7*, 19–22.

Dunn, K., and Griffith, J.D. (1980). The presence of RNA in a double helix inhibits its interaction with histone protein. Nucleic Acids Res. *8*, 555–566.

Edmond, V., Moysan, E., Khochbin, S., Matthias, P., Brambilla, C., Brambilla, E., Gazzeri, S., and Eymin, B. (2011). Acetylation and phosphorylation of SRSF2 control cell

fate decision in response to cisplatin. EMBO J. *30*, 510–523.

Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S., and Nudler, E. (2007). An Allosteric Path to Transcription Termination. Mol. Cell *28*, 991–1001.

Erdö, F., Trapp, T., Mies, G., and Hossmann, K.-A. (2004). Immunohistochemical analysis of protein expression after middle cerebral artery occlusion in mice. Acta Neuropathol. *107*, 127–136.

Felipe-abrio, I., Lafuente-barquero, J., García-rubio, M.L., and Aguilera, A. (2014). RNA polymerase II contributes to preventing transcription-mediated replication fork stalls. 1–16.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. Nucleic Acids Res. *42*.

Fordice, J., Kershaw, C., El-Naggar, A., and Goepfert, H. (1999). Adenoid cystic carcinoma of the head and neck: predictors of morbidity and mortality. Arch. Otolaryngol. Head. Neck Surg. *125*, 149–152.

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

Gallego-Paez, L.M., Tanaka, H., Bando, M., Takahashi, M., Nozaki, N., Nakato, R., Shirahige, K., and Hirota, T. (2014). Smc5/6-mediated regulation of replication progression contributes to chromosome assembly during mitosis in human cells. Mol. Biol. Cell *25*, 302–317.

Galvin, B.D., Denning, D.P., and Horvitz, H.R. (2011). SPK-1, an SR protein kinase, inhibits programmed cell death in Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A. *108*, 1998–2003.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1.

Ghosh, G., and Adams, J. a. (2011). Phosphorylation mechanism and structure of serine-arginine protein kinases. FEBS J. *278*, 587–597.

Giannakouros, T., Nikolakaki, E., Mylonis, I., and Georgatsou, E. (2011). Serine-arginine protein kinases: A small protein kinase family with a large cellular presence. FEBS J. *278*, 570–586.

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.

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.

Gorrini, C., Squatrito, M., Luise, C., Syed, N., Perna, D., Wark, L., Martinato, F., Sardella,

D., Verrecchia, A., Bennett, S., et al. (2007). Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. Nature *448*, 1063–1067.

Grabczyk, E., Mancuso, M., and Sammarco, M.C. (2007). A persistent RNA??DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. Nucleic Acids Res. *35*, 5351–5359.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: At the heart of the spliceosome. Rna *11*, 533–557.

Groh, M., and Gromak, N. (2014). Out of Balance: R-loops in Human Disease. PLoS Genet. *10*, e1004630.

Groh, M., Lufino, M.M.P., Wade-Martins, R., and Gromak, N. (2014). R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. PLoS Genet. *10*.

Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vítor, A.C., Desterro, J.M., Carmo-Fonseca, M., and de Almeida, S.F. (2015). Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. Elife *4*, e09214.

Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. Genes Dev. *29*, 849–861.

Haeusler, A.R., Donnelly, C.J., and Rothstein, J.D. (2016). The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease. Nat Rev Neurosci *17*, 383–395.

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.

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

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. Cell *144*, 646–674.

Hantsche, M., and Cramer, P. (2016). The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. Angew. Chemie Int. Ed. 2–12.

Harinarayanan, R., and Gowrishankar, J. (2003). Host factor titration by chromosomal R-loops as a mechanism for runaway plasmid replication in transcription termination-defective mutants of Escherichia coli. J. Mol. Biol. *332*, 31–46.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied

Roles for Each Phospho-CTD Residue. Cell Rep. *15*, 2147–2158.

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.

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

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

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

Herrera-Moyano, E., Mergui, X., García-Rubio, M.L., Barroso, S., and Aguilera, A. (2014). The yeast and human FACT chromatinreorganizing complexes solve R-loopmediated transcription-replication conflicts. Genes Dev. *28*, 735–748.

Hill, S.J., Rolland, T., Adelmant, G., Xia, X., Owen, M.S., Dricot, A., Zack, T.I., Sahni, N., Jacob, Y., Hao, T., et al. (2014). Systematic screening reveals a role for BRCA1 in the response to transcription-associated DNA damage. Genes Dev. *28*, 1957–1975.

Hill, S.J., Mordes, D.A., Cameron, L.A., Neuberg, D.S., Landini, S., Eggan, K., and Livingston, D.M. (2016). Two familial ALS proteins function in prevention/repair of transcription-associated DNA damage. Proc. Natl. Acad. Sci. U. S. A. 201611673.

Ho, A.S., Kannan, K., Roy, D.M., Morris, L.G.T., Ganly, I., Katabi, N., Ramaswami, D., Walsh, L. a, Eng, S., Huse, J.T., et al. (2013). The mutational landscape of adenoid cystic carcinoma. Nat. Genet. *45*, 791–798.

Hoeijmakers, J.H.J. (2015). DNA Damage, Aging, and Cancer. 1475–1485.

Hong, X., Cadwell, G.W., and Kogoma, T. (1995). Escherichia coli RecG and RecA proteins in R-loop formation. EMBO J. *14*, 2385–2392.

Hong, Y., Jang, S.W., and Ye, K. (2011). The N-terminal fragment from caspase-cleaved serine/arginine protein-specific kinase2 (SRPK2) translocates into the nucleus and promotes apoptosis. J. Biol. Chem. *286*, 777–786.

Hong, Y., Chan, C.B., Kwon, I.-S., Li, X., Song, M., Lee, H.-P., Liu, X., Sompol, P., Jin, P., Lee, H., et al. (2012). SRPK2 phosphorylates tau and mediates the cognitive defects in Alzheimer's disease. J. Neurosci. *32*, 17262–17272.

Huang, H.-S., Allen, J. a., Mabb, A.M., King, I.F., Miriyala, J., Taylor-Blake, B., Sciaky, N., Dutton, J.W., Lee, H.-M., Chen, X., et al. (2011). Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. Nature *481*, 185–189.

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

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071–1078.

Jackson, B.R., Noerenberg, M., and Whitehouse, A. (2014). A Novel Mechanism Inducing Genome Instability in Kaposi's Sarcoma-Associated Herpesvirus Infected Cells. PLoS Pathog. *10*.

Jády, B.E., Ketele, A., and Kiss, T. (2012). Human intron-encoded Alu RNAs are processed and packaged into Wdr79-associated nucleoplasmic box H/ACA RNPs. Genes Dev. *26*, 1897–1910.

Jang, S.W., Yang, S.J., Ehlén, Å., Dong, S., Khoury, H., Chen, J., Persson, J.L., and Ye, K. (2008). Serine/arginine protein-specific kinase 2 promotes leukemia cell proliferation by phosphorylating acinus and regulating cyclin A1. Cancer Res. *68*, 4559–4570.

Jang, S.W., Liu, X., Fu, H., Rees, H., Yepes, M., Levey, A., and Ye, K. (2009). Interaction of Akt-phosphorylated SRPK2 with 14-3-3 mediates cell cycle and cell death in neurons. J. Biol. Chem. *284*, 24512–24525.

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

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

Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D. (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell *153*, 855–868.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Kamachi, M., Le, T.M., Kim, S.J., Geiger, M.E., Anderson, P., and Utz, P.J. (2002). Human autoimmune sera as molecular probes for the identification of an autoantigen kinase signaling pathway. J. Exp. Med. *196*, 1213–1225.

Karakama, Y., Sakamoto, N., Itsui, Y., Nakagawa, M., Tasaka-Fujita, M., Nishimura-Sakurai, Y., Kakinuma, S., Oooka, M., Azuma, S., Tsuchiya, K., et al. (2010). Inhibition of hepatitis C virus replication by a specific inhibitor of serine-arginine-rich protein kinase. Antimicrob Agents Chemother *54*, 3179–3186.

Khobta, A., Ferri, F., Lotito, L., Montecucco, A., Rossi, R., and Capranico, G. (2006). Early effects of topoisomerase I inhibition on RNA polymerase II along transcribed genes in human cells. J. Mol. Biol. *357*, 127–138.

Khoronenkova, S. V, and Dianov, G.L. (2015). ATM prevents DSB formation by coordinating SSB repair and cell cycle progression. Proc. Natl. Acad. Sci. U. S. A. *112*, 3997–4002.

Kim, H.D., Choe, J., and Seo, Y.S. (1999). The sen1+ gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. Biochemistry *38*, 14697–14710.

Lahiry, P., Torkamani, A., Schork, N.J., and Hegele, R. a (2010). Kinase mutations in human disease: interpreting genotype-phenotype relationships. Nat. Rev. Genet. *11*, 60–74.

Landt, S., and Marinov, G. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome … 1813–1831.

de Lange, T. (2009). How telomeres solve the end-protection problem. Science (80-.). *326*, 948–952.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu. Rev. Genet. *34*, 77–137.

Lesnik, E. a, and Freier, S.M. (1995). Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. Biochemistry *34*, 10807–10815.

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

Liang, Y., Lin, S.-Y., Brunicardi, F.C., Goss, J., and Li, K. (2009). DNA damage response pathways in tumor suppression and cancer treatment. World J. Surg. *33*, 661–666.

Lim, Y.W., Sanz, L. a, Xu, X., Hartono, S.R., and Chédin, F. (2015). Genome-wide DNA hypomethylation and RNA:DNA hybrid accumulation in Aicardi–Goutières syndrome. Elife *4*, 1–21.

Lin, J.-C., Lin, C.-Y., Tarn, W.-Y., and Li, F.-Y. (2014). Elevated SRPK1 lessens apoptosis in breast cancer cells through RBM4-regulated splicing events. RNA *20*, 0–11.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.-D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nat. Struct. Mol. Biol. *15*, 819–826.

Lindahl, T., and Barnes, D.E. (2000). Repair of endogenous DNA damage. In Cold Spring Harbor Symposia on Quantitative Biology, pp. 127–133.

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

Linhart, H.G., Lin, H., Yamada, Y., Moran, E., Steine, E.J., Gokhale, S., Lo, G., Cantu, E., Ehrich, M., He, T., et al. (2007). Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing. Genes Dev. *21*, 3110–3122.

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, H., Hu, X., Zhu, Y., Jiang, G., and Chen, S. (201AD). Up-regulation of SRPK1 in

non-small cell lung cancer promotes the growth and migration of cancer cells. Tumor Biol. *37*, 7287–7293.

Liu, L.F., Duann, P., Lin, C.T., D'Arpa, P., and Wu, J. (1996). Mechanism of action of camptothecin. In Annals of the New York Academy of Sciences, pp. 44–49.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem. J. *417*, 15–27.

Maizels, N., and Gray, L.T. (2013). The G4 Genome. PLoS Genet. *9*.

Manley, J.L., and Tacke, R. (1996). SR proteins and splicing control. Genes Dev. *10*, 1569–1579.

Marinello, J., Chillemi, G., Bueno, S., Manzo, S.G., and Capranico, G. (2013). Antisense transcripts enhanced by camptothecin at divergent CpG-island promoters associated with bursts of topoisomerase I-DNA cleavage complex and R-loop formation. Nucleic Acids Res. *41*, 10110–10123.

Marinello, J., Bertoncini, S., Aloisi, I., Cristini, A., Malagoli Tagliazucchi, G., Forcato, M., Sordet, O., and Capranico, G. (2016). Dynamic Effects of Topoisomerase I Inhibition on R-Loops and Short Transcripts at Active Promoters. PLoS One *11*, e0147053.

Mathew, R., Hartmuth, K., Möhlmann, S., Urlaub, H., Ficner, R., and Lührmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. Nat. Struct. Mol. Biol. *15*, 435–443.

Matsuoka, S., Ballif, B.A., Smogorzewska, A., McDonald, E.R., Hurov, K.E., Luo, J., Bakalarski, C.E., Zhao, Z., Solimini, N., Lerenthal, Y., et al. (2007). ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science (80-.). *316*, 1160–1166.

Mendoza, O., Gueddouda, N.M., Boulé, J.-B., Bourdoncle, A., and Mergny, J.-L. (2015). A fluorescence-based helicase assay: application to the screening of G-quadruplex ligands. Nucleic Acids Res. *43*, e71.

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.

Misteli, T., Cáceres, J.F., Clement, J.Q., Krainer, A.R., Wilkinson, M.F., and Spector, D.L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. J. Cell Biol. *143*, 297–307.

Montecucco, A., and Biamonti, G. (2013). Pre-mRNA processing factors meet the DNA damage response. Front. Genet. *4*, 1–14.

Muhlmann, S., Mathew, R., Neumann, P., Schmitt, A., Luhrmann, R., and Ficner, R. (2014). Structural and functional analysis of the human spliceosomal DEAD-box helicase Prp28. Acta Crystallogr. Sect. D Biol. Crystallogr. *70*, 1622–1630.

Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N.A., Broin, P.Ó., 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 1–19.

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

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell *161*, 526–540.

Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., and Fischer, T. (2016). Transient RNA-DNA Hybrids are Required for Efficient Double-Strand Break Repair. Cell 1–13.

Oltean, S., and Bates, D.O. (2013). Hallmarks of alternative splicing in cancer. Oncogene *33*, 1–8.

Pandit, S., Zhou, Y., Shiue, L., Coutinho-Mansfield, G., Li, H., Qiu, J., Huang, J., Yeo, G.W., Ares, M., and Fu, X.D. (2013). Genome-wide Analysis Reveals SR Protein Cooperation and Competition in Regulated Splicing. Mol. Cell *50*, 223–235.

Papoutsopoulou, S., Nikolakaki, E., Chalepakis, G., Kruft, V., Chevaillier, P., and Giannakouros, T. (1999). SR protein-specific kinase 1 is highly expressed in testis and phosphorylates protamine 1. Nucleic Acids Res. *27*, 2972–2980.

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.

Paulsen, R.D., Sollier, J., Stork, C.T., and Cimprich, K.A. (2014). Short Article Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. 1–9.

Pefanis, E., and Basu, U. (2015). RNA Exosome Regulates AID DNA Mutator Activity in the B Cell Genome. In Advances in Immunology, pp. 257–308.

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

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

Pfister, S.X., Ahrabi, S., Zalmas, L.P., Sarkar, S., Aymard, F., Bachrati, C.Z., Helleday, T., Legube, G., LaThangue, N.B., Porter, A.C.G., et al. (2014). SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. Cell Rep. *7*, 2006–2018.

Porrua, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes. Trends Genet. *xx*, 1–15.

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.

Price, A.M., Görnemann, J., Guthrie, C., and Brow, D. a (2014). An unanticipated early function of DEAD-box ATPase Prp28 during commitment to splicing is modulated by U5 snRNP protein Prp8. RNA *20*, 46–60.

Proudfoot, N.J. (1989). How RNA polymerase II terminates transcription in higher eukaryotes. Trends Biochem. Sci. *14*, 105–110.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science (80-.). *352*, 1466–1468.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Radhakrishnan, A., Nanjappa, V., Raja, R., Sathe, G., Chavan, S., Nirujogi, R.S., Patil, A.H., Solanki, H., Renuse, S., Sahasrabuddhe, N.A., et al. (2016). Dysregulation of splicing proteins in head and neck squamous cell carcinoma. Cancer Biol. Ther. *17*, 219–229.

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

Reddy, K., Tam, M., Bowater, R.P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y.H., and Pearson, C.E. (2011). Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. Nucleic Acids Res. *39*, 1749–1762.

Ritchie, D.B., Schellenberg, M.J., Gesner, E.M., Raithatha, S. a, Stuart, D.T., and Macmillan, A.M. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. Nat. Struct. Mol. Biol. *15*, 1199–1205.

Rocak, S., and Linder, P. (2004). DEAD-box proteins: the driving forces behind RNA metabolism. Nat. Rev. Mol. Cell Biol. *5*, 232–241.

Rosonina, E., Kaneko, S., and Manley, J.L. (2006). Terminating the transcript: Breaking up is hard to do. Genes Dev. *20*, 1050–1056.

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

Salvi, J.S., and Mekhail, K. (2015). R-loops highlight the nucleus in ALS. Nucleus *6*, 23–29.

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

Santos-Pereira, J.M., Garcia-Rubio, M.L., Gonzalez-Aguilera, C., Luna, R., and Aguilera, a. (2014). A genome-wide function of THSC/TREX-2 at active genes prevents

transcription-replication collisions. Nucleic Acids Res. *42*, 12000–12014.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Ginno, P.A., Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., and Ginno, P.A. (2016). Prevalent , Dynamic , and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Mol. Cell 1–12.

Sapra, A.K., Änkö, M.L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E.M., and Neugebauer, K.M. (2009). SR Protein Family Members Display Diverse Activities in the Formation of Nascent and Mature mRNPs In Vivo. Mol. Cell *34*, 179–190.

Savage, K.I., Gorski, J.J., Barros, E.M., Irwin, G.W., Manti, L., Powell, A.J., Pellagatti, A., Lukashchuk, N., McCance, D.J., McCluggage, W.G., et al. (2014). Identification of a BRCA1-mRNA Splicing Complex Required for Efficient DNA Repair and Maintenance of Genomic Stability. Mol. Cell *54*, 445–459.

Schultz, S.J., and Champoux, J.J. (2008). RNase H activity: Structure, specificity, and function in reverse transcription. Virus Res. *134*, 86–103.

Shen, Z. (2011). Genomic instability and cancer: An introduction. J. Mol. Cell Biol. *3*, 1–3.

Skourti-stathaki, K., Kamieniarz-gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature.

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.

Sollier, J., and Cimprich, K. a. (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol. *25*, 514–522.

Sparks, J.L., Chon, H., Cerritelli, S.M., Kunkel, T.A., Johansson, E., Crouch, R.J., and Burgers, P.M. (2012). RNase H2-Initiated Ribonucleotide Excision Repair. Mol. Cell *47*, 980–986.

Spiers, A., Esseltine, D., Ruckdeschel, J., Davies, J., and Horton, J. (1996). Metastatic Adenoid Cystic Carcinoma of Salivary Glands: Case Reports and Review of the Literature. Cancer Control *3*, 336–342.

Spiro, R.H. (1997). Distant metastasis in adenoid cystic carcinoma of salivary origin. Am. J. Surg. *174*, 495–498.

Stojdl, D.F., and Bell, J.C. (1999). SR protein kinases: the splice of life. Biochem. Cell Biol. *77*, 293–298.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C.L., and Lührmann, R. (1997). The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase

with significant homology to the yeast splicing factor Prp28p. RNA *3*, 1313–1326.

Thomas, M., White, R.L., and Davis, R.W. (1976). Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc. Natl. Acad. Sci. U. S. A. *73*, 2294–2298.

Tijerina, P., Bhaskaran, H., and Russell, R. (2006). Nonspecific binding to structured RNA and preferential unwinding of an exposed helix by the CYT-19 protein, a DEAD-box RNA chaperone. Proc. Natl. Acad. Sci. U. S. A. *103*, 16698–16703.

Titov, D. V, Gilman, B., He, Q.-L., Bhat, S., Low, W.-K., Dang, Y., Smeaton, M., Demain, A.L., Miller, P.S., Kugel, J.F., et al. (2011). XPB, a subunit of TFIIH, is a target of the natural product triptolide. Nat. Chem. Biol. *7*, 182–188.

Tresini, M., Warmerdam, D.O., Kolovos, P., Snijder, L., Vrouwe, M.G., Demmers, J. a. a., van IJcken, W.F.J., Grosveld, F.G., Medema, R.H., Hoeijmakers, J.H.J., et al. (2015). The core spliceosome as target and effector of non-canonical ATM signalling. Nature.

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.

Twyffels, L., Gueydan, C., and Kruys, V. (2011). Shuttling SR proteins: More than splicing factors. FEBS J. *278*, 3246–3255.

Varjosalo, M., Keskitalo, S., VanDrogen, A., Nurkkala, H., Vichalkovski, A., Aebersold, R., and Gstaiger, M. (2013). The Protein Interaction Landscape of the Human CMGC Kinase Group. Cell Rep. *3*, 1306–1320.

Vivarelli, S., Lenzken, S.C., Ruepp, M.D., Ranzini, F., Maffioletti, A., Alvarez, R., Mühlemann, O., and Barabino, S.M.L. (2013). Paraquat Modulates Alternative Pre-mRNA Splicing by Modifying the Intracellular Distribution of SRPK2. PLoS One *8*.

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.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell *136*, 701–718.

Wang, H.Y., Lin, W., Dyck, J. a., Yeakley, J.M., Songyang, Z., Cantley, L.C., and Fu, X.D. (1998). SRPK2: A differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. J. Cell Biol. *140*, 737–750.

Wang, P., Zhou, Z., Hu, A., Pontede Albuquerque, C., Zhou, Y., Hong, L., Sierecki, E., Ajiro, M., Kruhlak, M., Harris, C., et al. (2014). Both decreased and increased SRPK1 levels promote cancer by interfering with PHLPP-mediated dephosphorylation of Akt. Mol. Cell *54*, 378–391.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5’ to 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.

Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. *3*, 1–2.

Wilson-Sali, T., and Hsieh, T.-S. (2002). Preferential cleavage of plasmid-based R-loops and D-loops by Drosophila topoisomerase IIIbeta. Proc. Natl. Acad. Sci. U. S. A. *99*, 7974–7979.

Xiang, S., Gapsys, V., Kim, H.Y., Bessonov, S., Hsiao, H.H., Möhlmann, S., Klaukien, V., Ficner, R., Becker, S., Urlaub, H., et al. (2013). Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. Structure *21*, 2162–2174.

Yang, Y., McBride, K.M., Hensley, S., Lu, Y., Chedin, F., and Bedford, M.T. (2014). Arginine Methylation Facilitates the Recruitment of TOP3B to Chromatin to Prevent R Loop Accumulation. Mol. Cell *53*, 484–497.

Yin, J., Park, G., Lee, J.E., Choi, E.Y., Park, J.Y., Kim, T.H., Park, N., Jin, X., Jung, J.E., Shin, D., et al. (2015). DEAD-box RNA helicase DDX23 modulates glioma malignancy via elevating MIR-21 biogenesis. Brain *138*, 2553–2570.

Yüce-Petronczki, O., and West, S.C. (2012). Senataxin, defective in the neurogenerative disorder AOA-2, lies at the interface of transcription and the DNA damage response. Mol. Cell. Biol. *33*, 406–417.

Zeller, P., Padeken, J., van Schendel, R., Kalck, V., Tijsterman, M., and Gasser, S.M. (2016). Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA hybrid-associated repeat instability. Nat. Genet. 1–13.

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

Zhong, X.Y., Ding, J.H., Adams, J. a., Ghosh, G., and Fu, X.D. (2009). Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones. Genes Dev. *23*, 482–495.

Zhou, Z., and Fu, X.D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122*, 191–207.

Zimmer, A.D., and Koshland, D. (2016). Differential roles of the RNases H in preventing chromosome instability. Proc. Natl. Acad. Sci. 201613448.

It is very simple to be happy, but it is very difficult to be simple.

Tagore

It's not what you look that matters, it's what you see.

Henry David Thoreau

Appendix C. Abbreviations

Abbreviations

ΔHel	Helicase truncation
µm	Microns
53BP1	p53-binding protein 1
ACC	Adenoid cystic carcinoma
ACTB	β-actin locus
AGS	Aicardi-Goutières Syndrome
AKT	Protein Kinase B
ALS	Amyotrophic lateral sclerosis
APOE	Apolipoprotein E locus
AQR	Aquarius helicase
ATP	Adenosine triphosphate
bp	Base pairs
C.elegans	*Caenorhabditis elegans*
CAN	Copy number alterations
CpG	Cytosine base followed immediately by a Guanine base
CTD-RNA Pol II	C-terminal domain of RNA polymerase II
DDX23	DEAD box helicase 23
DNA	Deoxyribonucleic acid
DNMT3B1	DNA cytosine-5 methyltransferase 3b isoform 1
DRIP/DRIP-seq	DNA-RNA immunoprecipitation/DNA-RNA immunoprecipitation sequencing
DSB(s)	DNA double strand break(s)
dsDNA	Double stranded DNA
dsRNA	Double stranded RNA
FACT	FAcilitates Chromatin Transcription complex
FMR1	Fragile X mental retardation 1 locus
G4-DNA	G-quadraplex DNA
GRO-seq	Global Run-on sequencing
H2A	Histone 2A
H2B	Histone 2B
H3	Histone 3
H3K36me3	Trimethylation of lysine 36 of histone H3
H3K9me3	Dimethylation of lysine 9 of histone H3
HeLa	Cervical adenocarcinoma cells
kb	Kilobases
Kda	Kilo Daltons
mRNA	Messenger RNA
ncRNA(s)	Non-protein coding RNA(s)
PAS	Poly-adenylation sites
PhosIA	Phospho inactive
PhosM	Phosphomimetic
RNA	Ribonucleic acid

RNA Pol I/II/III	RNA Polymerase I, RNA Polymerase II, RNA Polymerase III
RNAi	RNA interference
RNA-seq	RNA sequencing
RNP	Ribonucleoprotein
RPA	Replication protein A
RPKMs	Reads per kb per million mapped reads
S.cerevisiae	*Saccharomyces cerevisiae*
S.Pombe	Schizosaccharomyces pombe
Sen1/SETX	Senataxin helicase
Ser5p RNA Pol II	Serine 5 phosphorylated RNA polymerase II
SETD2	SET Domain Containing 2
siRNA(s)	small interfering RNA(s)
snRNA(s)	Small nuclear RNA(s)
snRNP(s)	Small nuclear ribonucloprotein particle(s)
SNRPN	Small Nuclear Ribonucleoprotein Polypeptide N locus
SR protein	Serine/arginine group of proteins
SRPK1	Serine/arginine protein kinase 1
SRPK2	Serine/arginine protein kinase 2
SRSF1 (or ASF/SF2)	Serine/arginine splicing factor 1
SSB(s)	DNA single strand break(s)
ssDNA	Single stranded DNA
THO/TREX	TRanscription EXport complex
THOC1	THO complex 1
Tip60	Histone acetyltransferase KAT5
TOP1	Human topoisomerase 1
TPMs	Transcripts per million
U1 snRNP	U1 Small nuclear ribonucloprotein particle(s)
U2-OS	Osteosarcoma cells
U4/U6.U5 tri-snRNP	U4/U6.U5 tri-Small nuclear ribonucloprotein particle(s)
U5 snRNP	U5 Small nuclear ribonucloprotein particle(s)
UV	Ultraviolet light
WT	Wild type
XRN2/Xrn2/Rat1	5'-3' exoribonuclease 2
γH2AX	Serine 139 phosphorylated histone 2A variant x
ChIP/ChIP-seq	Chromatin immunoprecipitaion/chromatin immunoprecipitation sequencing
RT-qPCR	Real time quantitative polymerase chain reaction

Hard writing makes easy reading.

Wallace Stegner

A wise man once said of his recently published book, "This is a celebration of my ego!"

Anonymous

Appendix D.
Articles

Articles related to this work

Sridhara SC, Carvalho S, Grosso AR, Gallego-Paez LM, Carmo-Fonseca M, de Almeida SF. Transcription dynamics prevents RNA-mediated genomic instability through SRPK2-dependent DDX23 phosphorylation. **Cell Rep. 2017** Jan 10;18(2):334-343.

Other relevant articles

- ✓ Bonnet A, Grosso AR, Elkaoutari A, Coleno E, Presle A, **Sridhara SC**, Janbon G, Géli V, deAlmeida SF, Palancade B. Introns protect eukaryotic genomes from transcription-associated genetic instability. Mol Cell. 2017 Aug 17;67(4):608-621.
- ✓ Carvalho S, Vitor A, **Sridhara SC**, Martins FB, Raposo AC, Desterro JM, Ferreira J, de Almeida SF. SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. **eLife, 2014** May 6. 10.7554/eLife.02482.
- ✓ Carvalho S, Raposo AC, Martins FB, Grosso AR, **Sridhara SC**, Rino J, Carmo-Fonseca M, de Almeida SF. Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. **Nucleic Acids Research, 2013** Mar 1;41(5):2881-93.

OPEN ACCESS
CellPress

Cell Reports
Report

Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation

Sreerama Chaitanya Sridhara,[1] Sílvia Carvalho,[1] Ana Rita Grosso,[1] Lina Marcela Gallego-Paez,[1] Maria Carmo-Fonseca,[1] and Sérgio Fernandes de Almeida[1,2,*]
[1]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1600-276 Lisboa, Portugal
[2]Lead Contact
*Correspondence: sergioalmeida@fm.ul.pt
http://dx.doi.org/10.1016/j.celrep.2016.12.050

SUMMARY

Genomic instability is frequently caused by nucleic acid structures termed R-loops that are formed during transcription. Despite their harmful potential, mechanisms that sense, signal, and suppress these structures remain elusive. Here, we report that oscillations in transcription dynamics are a major sensor of R-loops. We show that pausing of RNA polymerase II (RNA Pol II) initiates a signaling cascade whereby the serine/arginine protein kinase 2 (SRPK2) phosphorylates the DDX23 helicase, culminating in the suppression of R-loops. We show that in the absence of either SRPK2 or DDX23, accumulation of R-loops leads to massive genomic instability revealed by high levels of DNA double-strand breaks (DSBs). Importantly, we found *DDX23* mutations in several cancers and detected homozygous deletions of the entire *DDX23* locus in 10 (17%) adenoid cystic carcinoma (ACC) samples. Our results unravel molecular details of a link between transcription dynamics and RNA-mediated genomic instability that may play important roles in cancer development.

INTRODUCTION

R-loops are nucleic acid structures formed by an RNA:DNA hybrid and a displaced single-stranded DNA (Sollier and Cimprich, 2015). They arise essentially during RNA polymerase II (RNA Pol II) transcription when the nascent RNA molecule hybridizes with the template DNA strand. In addition to an important regulatory role in diverse cellular processes such as gene expression (Santos-Pereira and Aguilera, 2015) and DNA repair (Hatchi et al., 2015; Tresini et al., 2015), R-loops also pose great threats to genome stability and should therefore be kept under physiological levels (Skourti-Stathaki and Proudfoot, 2014). RNA processing factors are known to prevent R-loop formation and the resulting genomic instability in human cells (Hamperl and Cimprich, 2014; Santos-Pereira and Aguilera, 2015), although the signaling cascades and molecular switches that lead to their mobilization and activation are essentially unknown. The activity of several RNA processing factors is regulated through phosphorylation, which is in many cases driven by serine/arginine protein kinase 1 (SRPK1) and SRPK2 (Varjosalo et al., 2013; Zhou and Fu, 2013). However, the roles of both SRPK1 and SRPK2 in R-loop metabolism and genome stability have not yet been investigated. Herein, we identify a role for SRPK2 and its substrate DEAD box protein 23 (DDX23) in preventing genomic instability. We show that loss of each of these two proteins results in the accumulation of R-loops, which lead to DNA damage. DDX23 is a RNA helicase that is part of the spliceosomal U5 small nuclear ribonucleoprotein (U5 snRNP) (Teigelkamp et al., 1997). Importantly, we found that the role of DDX23 in suppressing R-loops does not require a functional U5 small nuclear ribonucleoprotein particle (snRNP), as depletion of either *PRP8* or *PRP6*, core components of the U5 snRNP (Achsel et al., 1998; Agafonov et al., 2016; Grainger and Beggs, 2005; Makarov et al., 2000), does not drive genomic instability. Moreover, we describe a pathway that employs RNA Pol II pausing to signal the location of R-loops and nucleate SRPK2-dependent DDX23 phosphorylation. This may constitute a genome caretaker mechanism that would operate to ward off against tumor-driving DNA damage episodes. Supporting this hypothesis, we observed that loss of *DDX23* is a prevalent feature of adenoid cystic carcinoma (ACC), an aggressive salivary gland cancer with very limited treatment options, mostly because of our incomplete understanding of its genomic foundations and molecular basis (Chae et al., 2015). Altogether, we show that oscillations in RNA Pol II dynamics are a molecular sensor of R-loops that prevents potentially catastrophic genomic instability events.

RESULTS

SRPK2 Is Necessary to Protect against the Accumulation of R-Loops

To gain additional insights into the biological function of SRPK1 and SRPK2 kinases, we depleted each of these kinases from human osteosarcoma (U-2OS) and HeLa cells by RNAi and measured the amount of resulting DNA damage. Following depletion of *SRPK2*, but not of *SRPK1*, there was a significant increase in nuclear phosphorylated H2AX (γH2AX) foci (here used

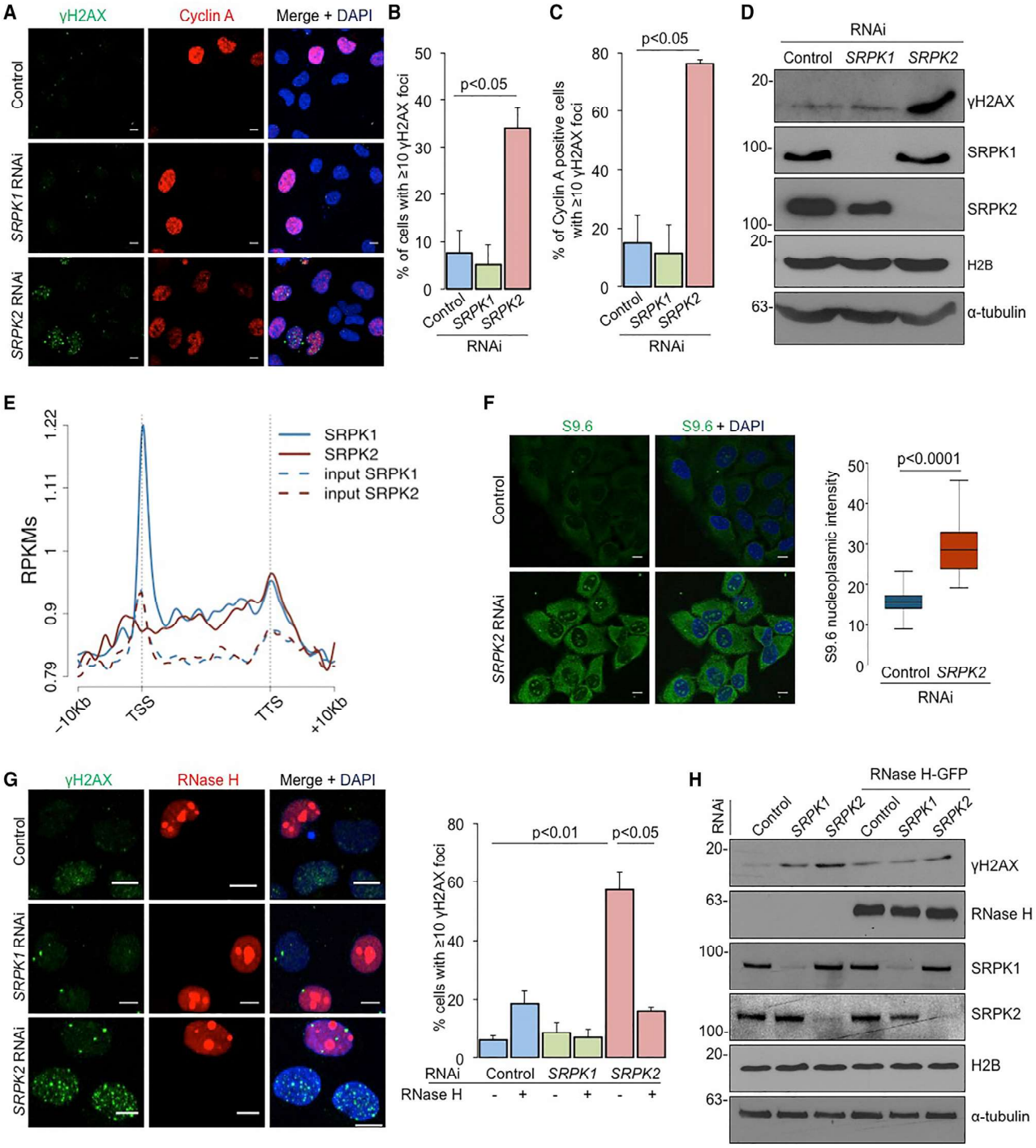


Figure 1. SRPK2 Prevents RNA-Dependent Genomic Instability

(A) γH2AX foci and cyclin A staining in control, *SRPK1*, and *SRPK2* RNAi-depleted U-2OS cells.

(B) Means and SDs of the percentage of cells with ten or more γH2AX foci. At least 100 cells from three independent experiments were scored.

(C) Same as (B) but for cyclin-A-positive cells only.

(D) Immunoblots showing γH2AX, SRPK1, and SRPK2 in U-2OS cells upon depletion of *SRPK1* or *SRPK2* by RNAi. Histone H2B and α-tubulin served as loading controls. Molecular weight markers (kilodaltons) are shown on the left. Data are representative of three independent experiments.

(E) Metagene analysis of SRPK1 and SRPK2 genome-wide average profiles. The average ChIP-seq and input signals (reads per kilobase of transcript per million mapped reads [RPKMs]) are shown for each kinase. The gene-body region was scaled to 200 windows using cubic spline interpolation, and ±10-kb gene-flanking regions were averaged in 100-bp windows.

(legend continued on next page)

as a proxy for DNA double-strand breaks [DSBs]), which was particularly prominent in replicating (cyclin-A-positive) cells (Figures 1A–1C). The augmented levels of γH2AX detected in whole extracts of U-2OS (Figure 1D) and HeLa (Figure S1A) cells further confirmed that SRPK2 activity is necessary to maintain genome stability.

Within the cell, SRPK1 and SRPK2 were detected both in the cytoplasm and in the nucleus, where they adopt a speckled distribution pattern (Figure S1B). In the nucleus, SRPK1 and SRPK2 were found both in the nucleoplasm and associated with chromatin (Figure S1C). The genome-wide analysis of the chromatin distribution of these two kinases revealed that they accumulate preferentially in coding regions of genes (Figure 1E). Examples of SRPK1 and SRPK2 distribution along individual genes are shown in Figure S2A. These data were validated by chromatin immunoprecipitation qPCR (ChIP-qPCR) (Figure S2B) with the same antibodies used for chromatin immunoprecipitation sequencing (ChIP-seq) (Figure S2C). While SRPK1 is enriched at the promoter region, SRPK2 is distributed more evenly across the entire gene (Figure 1E). Both kinases were preferentially detected at longer genes, but gene expression levels or GC content did not influence their chromatin binding (Figures S2D and S2E). Notably, however, we observed a strand asymmetry in the distribution of G and C residues (GC skew) in the vicinity of SRPK peaks (Figure S2D). Since GC skew is a predictor of R-loop formation (Ginno et al., 2013; Ginno et al., 2012), we wondered whether aberrant accumulation of R-loops drives the genomic instability observed in *SRPK2*-depleted cells. Immunofluorescence experiments performed with an antibody that specifically detects RNA:DNA hybrids (S9.6) (Boguslawski et al., 1986) revealed that depletion of *SRPK2* results in a robust accumulation of nucleoplasmic R-loops (Figure 1F). The cytoplasmic signal is likely to result from mitochondrial DNA replication as previously reported (Brown et al., 2008). Increased DNA damage and R-loop formation were also observed when *SRPK2* was depleted using two additional distinct small interfering RNAs (siRNAs) (Figures S3A–S3C). Importantly, digestion of R-loops by ribonuclease H1 (RNase H) overexpression led to a complete rescue of the DNA damage phenotype obtained upon *SRPK2* depletion as revealed by decreased nuclear foci (Figure 1G) and total cellular levels (Figure 1H) of γH2AX. Overexpression of RNase H did not change the total DNA content of the cells, indicating that the effect on DNA damage is not the result of a deregulated cell cycle progression (Figure S3D). Moreover, a similar phenotype rescue was obtained upon transcription inhibition with triptolide, a potent inhibitor of transcription initiation (Henriques et al., 2013), suggesting that active transcription is necessary to induce DNA damage in *SRPK2*-depleted cells (Figures S4A–S4C). The 70-min incubation with triptolide in our experimental setting did not produce any noticeable cell-cycle alterations (Figure S4D), suggesting that the effect on DNA damage is driven by transcription inhibition. Altogether, these data implicate SRPK2 in a mechanism that leads to co-transcriptional R-loop suppression and prevents RNA-mediated genomic instability.

A Phosphomimetic DDX23 Rescues the Genome Integrity in *SRPK2*-Depleted Cells

Both SRPKs act on proteins sharing a serine/arginine (RS)-rich domain. Among their substrates, SRSF1 (also known as ASF/SF2) is required to prevent R-loop-induced DNA DSBs (Li and Manley, 2005). We then investigated whether a lack of SRSF1 phosphorylation is responsible for the genomic instability observed in *SRPK2*-depleted cells. However, ectopic expression of a phosphomimetic version of SRSF1 (Cazalla et al., 2002) (SRSF1-PhosM) did not rescue the DNA damage phenotype in *SRPK2*-depleted cells (Figures 2A and S4E), suggesting that lack of SRSF1 phosphorylation is not causing the genomic instability observed in *SRPK2*-depleted cells.

The RS-domain-containing RNA helicase DDX23 is a member of the U5 snRNP that is phosphorylated by SRPK2, but not by SRPK1 (Laggerbauer et al., 1998; Mathew et al., 2008; Wahl et al., 2009). *SRPK2* knockdown results in hypophosphorylation of DDX23 and destabilizes its association with the U5 snRNP (Agafonov et al., 2016; Mathew et al., 2008). To address the role of DDX23 in the SRPK2-associated genomic instability phenotype, we replaced nine serine residues of the RS domain with nine aspartates to generate a phosphomimetic version of DDX23 (DDX23-PhosM). Notably, DDX23-PhosM was able to completely abolish the DNA damage induced by the *SRPK2* knockdown (Figures 2B, 2D, and S4F). Conversely, a non-phosphorylatable serine-to-alanine mutant of DDX23 (DDX23-PhosIA; Figure S4F) failed to prevent DNA damage following *SRPK2* depletion (Figure S4G). Moreover, a phosphomimetic DDX23 lacking the helicase domain (DDX23-PhosM-ΔHel) failed to rescue the DNA damage phenotype of *SRPK2*-depleted cells (Figures 2C, 2D, and S4F). The lack of DNA damage in control cells overexpressing the DDX23-PhosM-ΔHel (Figures 2C and 2D) suggests that this mutant protein does not have a dominant-negative effect on endogenous DDX23. Co-immunoprecipitation analyses revealed that DDX23-PhosM-ΔHel co-purified with histone H2A, suggesting that this mutant protein maintains its chromatin-binding capacity and only the helicase activity is disrupted (Figure S4H).

We next sought to investigate if the role of DDX23 in suppressing R-loops requires a functional U5 snRNP and if disruption of this complex causes genomic instability. To this end, we

(F) Immunofluorescence analysis of R-loops in control and *SRPK2* RNAi-depleted cells. The intensity of the nucleoplasmic staining is plotted. At least 100 cells from three independent experiments were scored. Statistical significance was determined using the Mann-Whitney test.
(G) γH2AX foci in control, *SRPK1*, and *SRPK2* RNAi-depleted cells transiently transfected with an RNase H-mCherry expression plasmid. Means and SDs of the percentage of cells with ten or more γH2AX foci are plotted. Note that the apparent increase in the percentage of control cells with γH2AX foci upon RNase H overexpression is not statistically significant. Data are from a minimum of 100 cells scored in four independent experiments.
(H) Immunoblots showing γH2AX, RNase H, SRPK1, and SRPK2 levels in *SRPK1* and *SRPK2* RNAi-depleted U-2OS cells that were transiently transfected with *RNase H-GFP*. Data are representative of three independent experiments performed.
Scale bars, 10 μm. Statistical significance was determined using two-tailed Student's t test.

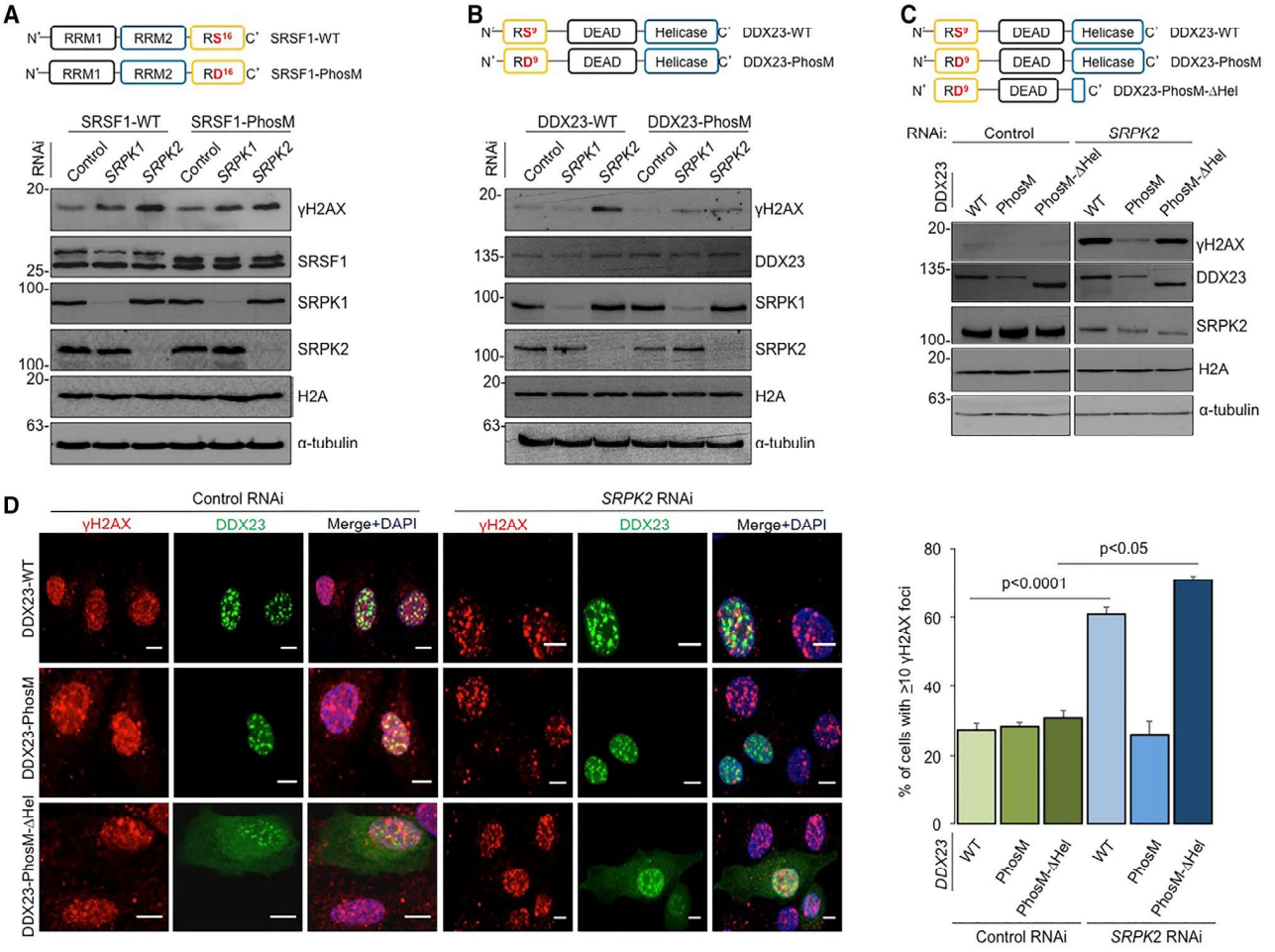


Figure 2. A Phosphomimetic Version of DDX23 Restores Genome Integrity in *SRPK2*-Depleted Cells

(A and B) Immunoblots showing γH2AX, SRSF1 (A), DDX23 (B), SRPK1, and SRPK2 levels in *SRPK1* and *SRPK2* RNAi-depleted U-2OS cells that were transiently transfected with *SRSF1-WT* or *SRSF1-PhosM* (A) or *DDX23-WT* or *DDX23-PhosM* (B). Histone H2A and α-tubulin served as loading controls.

(C) γH2AX, DDX23, and SRPK2 levels in *SRPK2* RNAi-depleted U-2OS cells that were transiently transfected with *DDX23-WT*, *DDX23-PhosM*, or *DDX23-PhosM-ΔHel*. Molecular weight markers (kilodaltons) are shown on the left. Data are representative of three independent experiments.

(D) γH2AX foci and DDX23 staining in control and *SRPK2* RNAi-depleted U-2OS cells upon transient transfection with *DDX23-WT*, *DDX23-PhosM*, or *DDX23-PhosM-ΔHel.* Means and SDs of the percentage of cells with ten or more γH2AX foci are plotted on the right-hand side. Data are from a minimum of 100 cells scored in three independent experiments. Scale bars shown in all immunofluorescence images represent 10 μm. Statistical significance was determined using two-tailed Student's t test.

performed RNAi against *PRP8*, a U5 snRNP scaffolding protein PRP8, which is at the core of U5 snRNP assembly and plays key roles during the catalytic activation of the spliceosome (Agafonov et al., 2016; Grainger and Beggs, 2005; Ritchie et al., 2008), did not induce any DNA damage (Figure 3A) and did not affect the ability of DDX23-PhosM to suppress DNA damage in *SRPK2*-depleted cells (Figures 3B and 3C). Similar data were obtained upon depletion of PRP6, another U5 snRNP component that is necessary for spliceosome activity (Makarov et al., 2000; Figures S5A–S5C). Altogether, these results show that phosphorylation of DDX23 is necessary and sufficient to restore the genome stability in *SRPK2*-deficient cells through a process that requires its RNA helicase activity, but not a functional U5 snRNP or spliceosome activity.

Loss of *DDX23* Drives RNA-Dependent Genomic Instability and Is Frequent in Adenoid Cystic Carcinoma

To further explore the role of DDX23, we measured R-loop formation upon *DDX23* knockdown (Figure 4A). In agreement with a role in RNA-mediated genome stability, depletion of *DDX23* caused a strong accumulation of R-loops (Figures 4A and S5D). These were further processed into DNA DSBs as revealed by the increased levels of γH2AX (Figure 4B) and by the concomitant accumulation of DSB-associated chromosomal aberrations observed in metaphase spreads of *DDX23* (and *SRPK2*) depleted cells (Figure 4C). Moreover, phosphorylation of 53BP1 and RPA proteins, which takes place during the repair of DNA DSBs, was also detected upon depletion of *DDX23* and *SRPK2* (Figure S5E). Increased DNA damage and R-loop

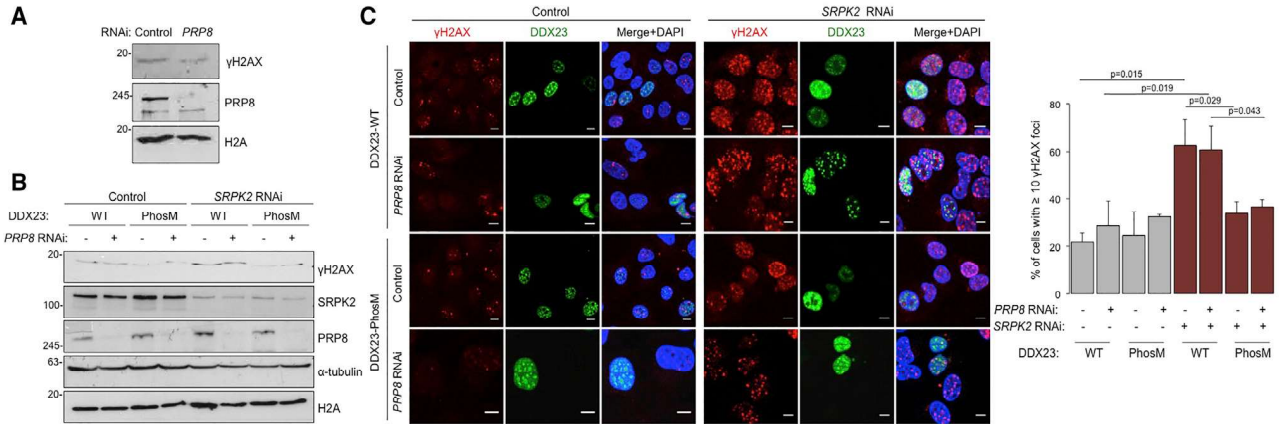


Figure 3. U5 snRNP and Spliceosome Activity Are Not Necessary for the Suppression of DNA Damage by DDX23

(A) γH2AX and PRP8 levels in *PRP8* RNAi-depleted U-2OS cells.

(B) γH2AX, SRPK2, and PRP8 levels in U-2OS cells following RNAi-depletion of *SRPK2* and *PRP8* and transiently transfection of *DDX23-WT* or *DDX23-PhosM*, as detailed in the figure. Molecular weight markers (kilodaltons) are shown on the left. α-Tubulin and histone H2A levels served as loading controls. Data are representative of three independent experiments.

(C) γH2AX foci and DDX23 staining in control, *SRPK2*, and *PRP8* RNAi U-2OS cells upon transient transfection with *DDX23-WT* and *DDX23-PhosM*. Means and SDs of the percentage of cells with ten or more γH2AX foci are plotted on the right-hand side. Data are from a minimum of 100 cells scored in three independent experiments. Scale bars shown in all immunofluorescence images represent 10 μm. Statistical significance was determined using two-tailed Student's t test.

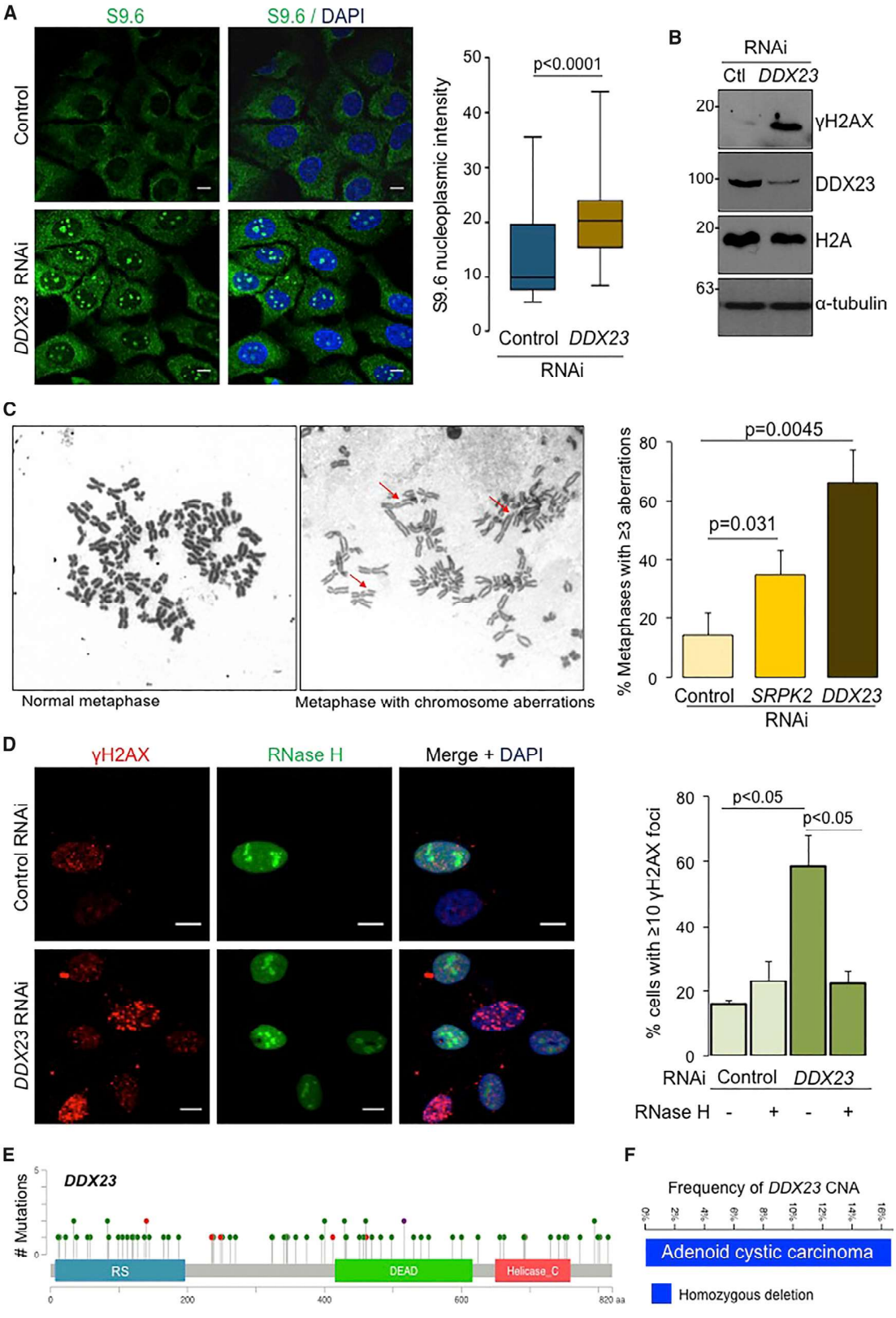
formation were also observed upon *DDX23* depletion with two additional distinct siRNAs (Figures S3A–S3C). Notably, the genomic instability phenotype of *DDX23*-depleted cells was rescued by overexpression of RNase H (Figure 4D), suggesting that R-loops are the source of the DNA damage observed in DDX23-deficient cells.

In agreement with the view that phosphorylation of DDX23 by SRPK2 constitutes a genome caretaker mechanism to ward off against tumorigenic processes, our pan-cancer data analysis revealed that 45% (24 out of 53) of all *DDX23* point mutations found in cancer samples are significant enriched in the RS domain (Fisher's exact test $p < 0.05$; Figure 4E). Moreover, 83% of these mutations are predicted to be potential deleterious variants (Table S1). Within this domain, 63% of the mutations affected either an arginine (R) or a serine (S) residue. Moreover, *DDX23* copy-number alterations were also observed in several cancers, such as ACC. Our analyses detected homozygous deletions of the *DDX23* locus in 17% (10 out of 60) of ACC samples (Figure 4F). Interestingly, the marked genomic instability that follows deletion of *DDX23* may explain the recurrent high-level losses in 12q13—the genomic housing of *DDX23*—previously observed in ACC (Ho et al., 2013).

RNA Pol II Pausing Nucleates SRPK2 and DDX23 at R-Loop-Containing Sites

We then inspected whether DDX23 is directly recruited to specific R-loop-containing chromatin loci. To induce R-loops, we depleted the *THOC1* component of the THO/TREX, a nuclear complex that functions in the biogenesis of messenger ribonucleoproteins and prevents R-loops in active genes (Domínguez-Sánchez et al., 2011; Huertas and Aguilera, 2003; Santos-Pereira et al., 2014). As previously described (Bhatia et al., 2014), RNA:DNA immunoprecipitation (DRIP) revealed that *THOC1* knockdown leads to accumulation of R-loops in the *APOE* and *ACTB* genes, but not in *SNRPN* (Figure 5A). In agreement with a role in the suppression of R-loops, *DDX23* knockdown also resulted in a prominent accumulation of these structures in *APOE* and *ACTB* (Figure 5A). We then explored if DDX23 is recruited to the chromatin region where R-loops formed upon *THOC1* knockdown. In fact, chromatin immunoprecipitation (ChIP) experiments showed a robust accumulation of DDX23 in R-loop-containing *APOE* and *ACTB* loci, but not in the R-loop-free *SNRPN* locus (Figure 5B). Importantly, suppression of R-loops by overexpression of RNase H significantly reduced the recruitment of DDX23 to *APOE* and *ACTB* loci in *THOC1*-depleted cells (Figure 5B). In contrast, this recruitment was not affected by *SRPK2* depletion (Figure S5F), suggesting that DDX23 phosphorylation is required for the protein activity (namely to resolve co-transcriptional R-loops), but not to its chromatin binding. These results reveal that DDX23 accumulates at chromatin regions containing R-loops and is necessary to suppress locus-specific R-loop formation.

Co-transcriptional R-loops impact RNA Pol II transcription dynamics (Huertas and Aguilera, 2003; Santos-Pereira and Aguilera, 2015; Tous and Aguilera, 2007). For instance, they are involved in promoter-proximal pausing and promote efficient transcription termination by slowing down RNA Pol II, facilitating the timely recruitment of termination factors (Mischo et al., 2011; Skourti-Stathaki et al., 2011, 2014; Zhao et al., 2016). We then sought to investigate whether changes in transcription dynamics reach back to signal the suppression of deleterious R-loops in a mechanism involving SRPK2 and DDX23. We first assessed the impact of the aberrant formation of R-loops on RNA Pol II transcription. In agreement with the view that R-loops slow down



(legend on next page)

transcription, we detected a significant accumulation of RNA Pol II in the *APOE* and *ACTB* loci where R-loops had been detected, but not in the *SNRPN* gene, upon *THOC1* knockdown (Figure 5C). *THOC1* knockdown did not cause significant changes in the expression levels of *APOE* and *ACTB* (Figure S5G), suggesting that the observed RNA Pol II accumulation is not the result of overall changes in transcription rate, but rather a localized pausing event. Indeed, the RNA Pol II accumulated at the *APOE* and *ACTB* loci carried a serine 5-phosphorylation (Ser5P) pattern (Figure 5D), which is suggestive of paused transcription complexes (Buratowski, 2009). Notably, the stalled RNA Pol II complexes were released upon suppression of R-loops by RNase H overexpression (Figures 5C and 5D). Given the intimate link with R-loops, we reasoned that paused RNA Pol II complexes would constitute ideal molecular sensors to signal the location of intragenic R-loops and initiate a molecular pathway toward their suppression. According to our model, paused RNA Pol II would nucleate SRPK2-dependent phosphorylation of DDX23. In agreement with our view, co-immunoprecipitation experiments revealed that the accumulation of R-loops upon *THOC1* knockdown increased Ser5P RNA Pol II levels and promoted the interaction of RNA Pol II with SRPK2, but not with SRPK1 (Figure 5E). Moreover, the analysis of ChIP-seq data revealed that SRPK2 is enriched at sites of RNA Pol II pausing (Figure 5F). Finally, we observed that SRPK2 coincides with DDX23 at *APOE* and *ACTB* (Figure S5H). Altogether, these data strengthen our proposed model whereby changes in transcription dynamics initiate a molecular pathway leading to SRPK2-dependent DDX23 phosphorylation and suppression of R-loops.

DISCUSSION

R-loops are natural byproducts of transcription that play important physiological roles in diverse cellular processes. However, unscheduled R-loop formation threatens genome stability, and dedicated cellular mechanisms must be on call to react promptly in such cases. Here, we describe one such mechanism involving RNA Pol II as an R-loop sensor, SRPK2 as a signal transducer, and DDX23 as the molecular effector against R-loops. We detect DDX23 accumulation at chromatin loci where R-loops form, which are also co-occupied by Ser5P-RNA Pol II and SRPK2. Given the association between Ser5P-RNA Pol II and transcription pausing (Alexander et al., 2010; Nojima et al., 2015), we suggest that the concurrent chromatin accumulation of SRPK2 and DDX23 follows R-loop-dependent RNA Pol II pausing. While we cannot exclude that SRPK2 accumulates on transcription pause sites at least partially due to their general role in phosphorylating mRNA processing factors, its increased association with RNA Pol II observed in *THOC1*-depleted cells suggests a specific recruitment to chromatin in response to R-loops accumulation. We reason that the accumulation of Ser5P RNA Pol II—predominant near the transcription start site (Heidemann et al., 2013)—within the gene body challenges the expected phosphorylation pattern of elongating polymerases and signals the recruitment of both SRPK2 and DDX23. A similar mechanism linking Ser5P RNA Pol II and the recruitment of RNA-processing factors may explain the splicing-dependent transcription pausing (Harlen et al., 2016; Nojima et al., 2015). What are the differences, if any, between an R-loop-dependent and a splicing-dependent RNA Pol II pausing, and how does transcription pausing signal the nucleation of distinct molecular machineries? These are important questions that emerge from our model. Importantly, we should also consider the hypothesis that R-loops may form in response to the splicing-dependent RNA Pol II pausing, as was recently shown in the case of transcription-blocking DNA lesions (Tresini et al., 2015). In this scenario, the SRPK2-DDX23 axis described here would be a good candidate to suppress such opportunistic R-loops. Further studies aimed at clarifying these open questions will certainly provide valuable insights into the mechanistic coupling between transcription, pre-mRNA processing and genomic instability.

Altogether, our data allow us to propose a model whereby R-loops trigger RNA Pol II pausing, which in turn nucleates SRPK2-dependent DDX23 phosphorylation, preventing RNA-dependent genomic instability. Our model implies that DDX23 activation follows the accumulation of R-loops, placing this helicase as a suppressor rather than a preventer of R-loops. This separates it from the role of several other RNA processing factors such as SRSF1 or THOC1 that prevent the formation of co-transcriptional RNA:DNA hybrids. Instead, DDX23 would fall within the family of R-loop suppressors, on which the helicase senataxin (SETX) occupies a prominent position. Notably, while SETX resolves RNA:DNA hybrids to promote transcription

Figure 4. Loss of *DDX23* Drives RNA-Dependent Genomic Instability and Is Frequent in Adenoid Cystic Carcinoma

(A) Immunofluorescence analysis of R-loops 48 hr after control or *DDX23* depletion in U-2OS cells. The intensity of the nucleoplasmic staining is plotted. At least 100 cells from three independent experiments were scored. Statistical significance was determined using the Mann-Whitney test.

(B) Immunoblots showing γH2AX and DDX23 levels in control and *DDX23*-depleted U-2OS cells. Histone H2A and α-tubulin served as loading controls. Molecular weight markers (kilodaltons) are shown on the left. Data are representative of three independent experiments.

(C) Metaphase spreads of cells transfected with siRNAs against *SRPK2* and *DDX23*. Two representative metaphases (one normal and one with chromosome aberrations) are shown. Red arrows point to DSB-associated chromosomal aberrations. The percentage of metaphases with three or more chromosome aberrations is plotted. Approximately 80 spreads were scored per condition. Means and SDs are shown. Statistical significance was determined using two-tailed Student's t test.

(D) γH2AX foci in control and *DDX23* RNAi-depleted U-2OS cells transiently transfected with an RNase H-GFP expression plasmid. Means and SDs of the percentage of cells with ten or more γH2AX foci are plotted. Data are from a minimum of 100 cells scored in three independent experiments. Scale bars on immunofluorescence images represent 10 μm. Statistical significance was determined using two-tailed Student's t test.

(E) Schematic representation of point mutations found in the *DDX23* locus in cancers represented in the cBioPortal (Cerami et al., 2012; Gao et al., 2013) (117 studies with mutation data). The RS, DEAD, and helicase domains of DDX23 are shown.

(F) Frequency of copy-number alterations (CNAs) of the *DDX23* locus in ACC samples. Data are from the Memorial Sloan Kettering Cancer Center (MSKCC) cBioPortal for Cancer Genomics.

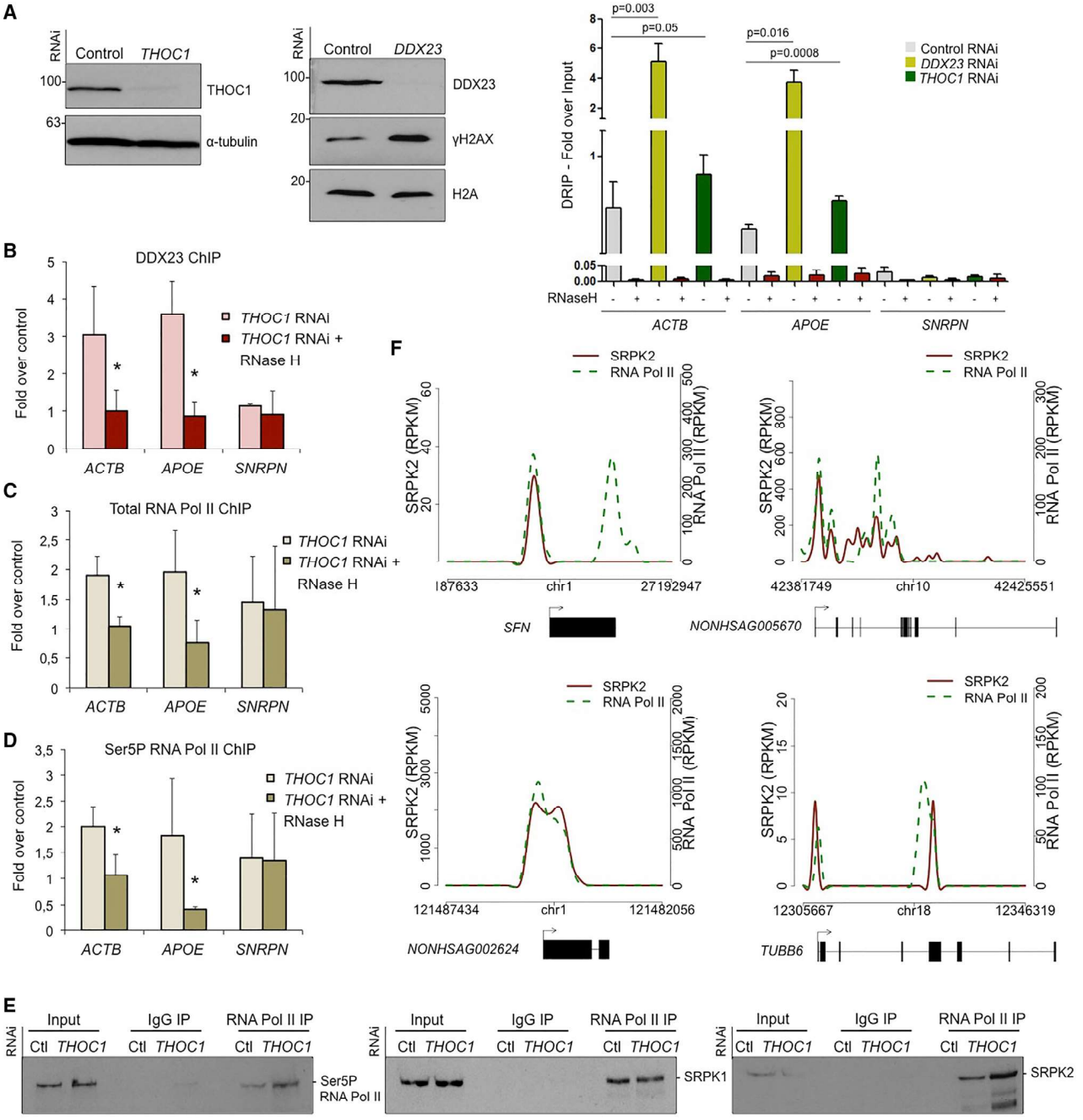


Figure 5. RNA Pol II Pausing Nucleates SRPK2 and DDX23 at R-Loop-Containing Sites

(A) Immunoblots of THOC1, DDX23, and γH2AX in U2-OS cells depleted of *THOC1* and *DDX23*. α-Tubulin and H2A served as loading controls. Molecular weight markers (kilodaltons) are shown on the left. RNA:DNA immunoprecipitations (DRIP) in *DDX23* and *THOC1* RNAi-depleted cells are shown in the right-hand side. Data are presented as fold increase over the input sample. Means and SDs from three independent DRIP experiments are plotted. Statistical significance was determined using two-tailed Student's t test.

(B–D) ChIP analysis of DDX23 (B), RNA Pol II (C), and Ser5P-RNA Pol II (D) in *THOC1* RNAi-depleted cells with (+ RNase H) or without transient transfection of *RNase H*. Data were normalized against the ChIP values obtained in control (i.e., upon RNAi using control siRNAs) cells. Means and SDs are from three independent ChIP experiments. *$p < 0.05$ (two-tailed Student's t test).

(E) Nuclear immunoprecipitations of RNA Pol II on control and *THOC1* RNAi-depleted U-2OS cells. Purified complexes were resolved by SDS-PAGE and blotted with antibodies against Ser5P-RNA Pol II, SRPK1, or SRPK2. The input lane represents total cell lysates, and "IgG IP" denotes the negative control immunoprecipitation. Data are representative of three independent experiments.

(F) Individual profiles with 20-bp windows of RNA Pol II and SRPK2 distribution (RPKMs) along four different loci.

termination (Skourti-Stathaki et al., 2011), we suggest that DDX23 acts to release RNA Pol II from R-loop-mediated pausing throughout the gene body during transcription elongation.

Importantly, our findings may illustrate a tumor suppressor function that employs oscillations in RNA Pol II transcription dynamics as a molecular sensor of potentially catastrophic and tumor-driving genomic instability events. The observation that *DDX23* is frequently lost in ACC unravels a candidate for further studies aimed at understanding the molecular basis of this cancer, a task that has been complicated by the lack of verified ACC cell lines (Phuchareon et al., 2009). Treatment of ACC is confined to surgery and radiation, while chemotherapy has been of limited palliative benefit in patients with advanced disease (Adelstein et al., 2012). Our study provides insights into the molecular framework of ACC that may set the ground for new molecularly targeted therapies.

EXPERIMENTAL PROCEDURES

Cells

Human osteosarcoma (U-2OS) and HeLa cells were grown as monolayers in DMEM (Invitrogen), supplemented with 10% (v/v) fetal bovine serum (FBS), 1% (v/v) nonessential amino acids, 1% (v/v) L-glutamine, and 100 U/mL penicillin-streptomycin and maintained at 37°C in a humidified atmosphere with 5% CO_2. To inhibit transcription, cells were treated for 70 min with 500 nM triptolide (Santa Cruz Biotechnology).

Immunofluorescence

R-loops were detected with the S9.6 antibody (ENH001, Kerafast) following cell fixation and permeabilization with 100% ice-cold methanol and acetone for 10 min and 1 min on ice, respectively. Incubation with primary antibodies was followed by incubation with fluorochrome-conjugated antibodies Cy3 (Jackson ImmunoResearch) and Dy488 (Bethyl Laboratories). All the washing steps were done with PBS containing 0.05% (vol/vol) Tween 20.

DRIP

RNA:DNA hybrids were immunoprecipitated as described earlier (Ginno et al., 2012), with the modifications detailed in Supplemental Experimental Procedures.

Western blot, immunoprecipitation, DRIP, ChIP, immunofluorescence, analyses of metaphase spreads, nucleic acids quantification, bioinformatics analyses, and statistical methods are described in detail in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the SRPK1 and SRPK2 ChIP-seq data reported in this paper is GEO: GSE76777.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.12.050.

AUTHOR CONTRIBUTIONS

S.F.d.A. supervised the project and wrote the manuscript with the help of S.C.S. S.C.S. and S.C. performed the wet-lab experiments. L.M.G.-P. analyzed chromosomal aberrations in metaphase spreads. A.R.G. did all computational analyses of RNA-sequencing and ChIP-seq data. S.F.d.A., S.C.S., S.C, A.R.G., and M.C.-F. suggested experiments and analyzed and discussed data. All authors read and edited the manuscript.

ACKNOWLEDGMENTS

We thank all Salmeida lab members for support, technical advice, and valuable discussions. We thank Dr. Frederic Chedin and our colleagues Robert Martin, Ines Pankonien, Rosana Rocha, Ana Nascimento, and the Flow Cytometry and the Bioimaging iMM facilities staff for technical assistance. We are grateful to Drs. Robert J. Crouch, Patrick Calsou, Javier Caceres, and Ralf Ficner for providing RNaseH-GFP, RNaseH-mCherry, SRSF1 WT/RD, and DDX23 expression plasmids, respectively. This work was supported by grants from Fundação para a Ciência e Tecnologia (FCT), Portugal (LISBOA-01-0145-FEDER-007391 and PTDC/BIM-ONC/0016-2014). S.C.S. is supported by RNPnet, a Marie Curie Initial Training Network (PITN-GA-2011-289007). A.R.G. is the recipient of a FCT Investigator award (IF/00510/2014).

Received: May 13, 2016
Revised: November 13, 2016
Accepted: December 14, 2016
Published: January 10, 2017

REFERENCES

Achsel, T., Ahrens, K., Brahms, H., Teigelkamp, S., and Lührmann, R. (1998). The human U5-220kD protein (hPrp8) forms a stable RNA-free complex with several U5-specific proteins, including an RNA unwindase, a homologue of ribosomal elongation factor EF-2, and a novel WD-40 protein. Mol. Cell. Biol. *18*, 6756–6766.

Adelstein, D.J., Koyfman, S.A., El-Naggar, A.K., and Hanna, E.Y. (2012). Biology and management of salivary gland cancers. Semin. Radiat. Oncol. *22*, 245–253.

Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. Science *351*, 1416–1420.

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

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

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.

Brown, T.A., Tkachuk, A.N., and Clayton, D.A. (2008). Native R-loops persist throughout the mouse mitochondrial DNA genome. J. Biol. Chem. *283*, 36743–36751.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Cazalla, D., Zhu, J., Manche, L., Huber, E., Krainer, A.R., and Cáceres, J.F. (2002). Nuclear export and retention signals in the RS domain of SR proteins. Mol. Cell. Biol. *22*, 6871–6882.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404.

Chae, Y.K., Chung, S.Y., Davis, A.A., Carneiro, B.A., Chandra, S., Kaplan, J., Kalyan, A., and Giles, F.J. (2015). Adenoid cystic carcinoma: current therapy and potential therapeutic advances based on genomic profiling. Oncotarget *6*, 37117–37134.

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

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1.

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.

Grainger, R.J., and Beggs, J.D. (2005). Prp8 protein: at the heart of the spliceosome. RNA *11*, 533–557.

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

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. Cell Rep. *15*, 2147–2158.

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.

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

Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2013). Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. Mol. Cell *52*, 517–528.

Ho, A.S., Kannan, K., Roy, D.M., Morris, L.G., Ganly, I., Katabi, N., Ramaswami, D., Walsh, L.A., Eng, S., Huse, J.T., et al. (2013). The mutational landscape of adenoid cystic carcinoma. Nat. Genet. *45*, 791–798.

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

Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. Proc. Natl. Acad. Sci. USA *95*, 4188–4192.

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

Makarov, E.M., Makarova, O.V., Achsel, T., and Lührmann, R. (2000). The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein-protein interactions. J. Mol. Biol. *298*, 567–575.

Mathew, R., Hartmuth, K., Möhlmann, S., Urlaub, H., Ficner, R., and Lührmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. Nat. Struct. Mol. Biol. *15*, 435–443.

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.

Nojima, T., Gomes, T., Grosso, A.R., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell *161*, 526–540.

Phuchareon, J., Ohta, Y., Woo, J.M., Eisele, D.W., and Tetsu, O. (2009). Genetic profiling reveals cross-contamination and misidentification of 6 adenoid cystic carcinoma cell lines: ACC2, ACC3, ACCM, ACCNS, ACCS and CAC2. PLoS ONE *4*, e6040.

Ritchie, D.B., Schellenberg, M.J., Gesner, E.M., Raithatha, S.A., Stuart, D.T., and Macmillan, A.M. (2008). Structural elucidation of a PRP8 core domain from the heart of the spliceosome. Nat. Struct. Mol. Biol. *15*, 1199–1205.

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

Santos-Pereira, J.M., García-Rubio, M.L., González-Aguilera, C., Luna, R., and Aguilera, A. (2014). A genome-wide function of THSC/TREX-2 at active genes prevents transcription-replication collisions. Nucleic Acids Res. *42*, 12000–12014.

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.

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.

Sollier, J., and Cimprich, K.A. (2015). Breaking bad: R-loops and genome integrity. Trends Cell Biol. *25*, 514–522.

Teigelkamp, S., Mundt, C., Achsel, T., Will, C.L., and Lührmann, R. (1997). The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. RNA *3*, 1313–1326.

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

Tresini, M., Warmerdam, D.O., Kolovos, P., Snijder, L., Vrouwe, M.G., Demmers, J.A., van IJcken, W.F., Grosveld, F.G., Medema, R.H., Hoeijmakers, J.H., et al. (2015). The core spliceosome as target and effector of non-canonical ATM signalling. Nature *523*, 53–58.

Varjosalo, M., Keskitalo, S., Van Drogen, A., Nurkkala, H., Vichalkovski, A., Aebersold, R., and Gstaiger, M. (2013). The protein interaction landscape of the human CMGC kinase group. Cell Rep. *3*, 1306–1320.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

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.

Zhou, Z., and Fu, X.D. (2013). Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma *122*, 191–207.

I meant what I said. If I could have said it any differently, I would have!

Eliot

Here is a chemical reaction; four years of life, turmoil, sweat, passion, hardships and love; boiled in a vessel with flesh, bones, muscle and so called soul, chilled over time; yielding not even four kilograms of bound paper! Here it is!

Anonymous

As for the thesis itself — a slim, red volume with gold lettering — it's not something she feels sentimental about. "I've met people who, they cry when they give away their kids' baby clothes, but I was never one of those — and I think I felt the same way about the thesis." She's more inclined to look forward. "In exoplanets, the best planet, the best discovery, is the next discovery."

Sara Seager

All good things must come to an end, and when the time comes to make my exit, I hope I can do so with good grace and humor, but there is time yet, and many small moments to savor.

Ruskin Bond

The End.

For a long time
it just sat there.

Back cover design: The image is from “The missing piece meets the big O” by Shel Silverstein © 1981. Harper Collin’s Pubslishers and brainpickings.org (Maria Popova). The wedge in the picture depicts paused RNA Pol II on DNA template.