UC Davis
UC Davis Electronic Theses and Dissertations

Title

SF3B1-targeted Splicing Inhibition Triggers Global Alterations in Transcriptional Dynamics and R-Loop Metabolism

Permalink

https://escholarship.org/uc/item/1q60s0tp

Author

Castillo Guzman, Daisy

Publication Date

2022

Peer reviewed|Thesis/dissertation

SF3B1-targeted Splicing Inhibition Triggers Global Alterations in
Transcriptional Dynamics and R-Loop Metabolism

By

DAISY CASTILLO GUZMAN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry, Molecular, Cellular, Developmental Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____
Frédéric Chédin, Chair

_____
Jacqueline Barlow

_____
Wolf-Dietrich Heyer

_____
Ben Montpetit

Committee in Charge

2022

Table of Contents

ACKNOWLEDGEMENTS

I am extremely grateful to my mentor Brian Haley. I still recall the first conversation we had together. It was a phone interview for a summer internship at Oregon State University. I was finishing up my last semester at Santa Rosa Junior College. Brian asked me questions about my research experience and I was honest with him in that I had zero research experience outside of my courses. Despite this, Brian said that as long as I was willing to put in the hard work and have fun that summer, I could work with him. As I look back, I now realize how much of a gamble that was for him, to undertake a student with no research experience and work with them for 10 weeks. This research experience was truly life-changing. I learned how fun it would be to pursue a research question and dedicate so much time tackling different experiments in pursuit of testing that question. It is because of the mentorship I received from Brian that I decided to attend graduate school in the first place. Thank you, Brian. I do not think you realize the impact you have had on my academic trajectory. I hope that in the future I can be a mentor like you and remind myself to take chances on students with limited research experience and help them grow and see all their potential.

I would like to thank the members of the Chédin lab. I had the pleasure of working with the following individuals during my graduate training: Lionel Sanz, Stella Hartono Shaheen Sulthana, Eriko Ijiro, Maika Malig, Robert Stolz, John Smolka, Talysa Viera, Meghan Frederick, Ethan Holleman, Tadas Sereiva, Aidan Du, Saloni Dhopte, Rachel Larson, Catherine Nugent, Miriam Rivkin, Kelly Chau, Kimberley Berg, Tommer Schwarz, Jenna Giafaglione, and Michelle Chui. I would like to specially thank two individuals who were instrumental in my graduate training, Lionel Sanz and Stella Hartono. I want to thank Lionel for all the support and mentorship he provided throughout my graduate training. At the beginning, I constantly asked him questions and he never once hesitated to answer them. I want to thank Stella, whom I worked extensively with on my dissertation project. I had so much fun working alongside her and learning from her. I will always have found memories of us working together to generate figures for our manuscript at the height of the pandemic! Lionel and Stella, thank you for being my mentors in the lab. I appreciate all the support you both have given me throughout the years.

I thank all my mentors for their advice. Wolf-Dietrich, Ben Montpetit, and Jackie Barlow, who have provided insightful feedback as part of my dissertation committee. They have all been very supportive

and encouraging mentors. Thank you for your comments and advice throughout the years. I appreciate the time you took to meet with me.

Lastly, I would like to express my sincere gratitude to my advisor, Frédéric Chédin. Thank you for giving me the opportunity to work with you and all the members of your lab. It has been fun growing as a scientist in your lab and learning about the world of R-loops. Thank you for always supporting me in my research endeavors and in providing a safe environment to troubleshoot my way to becoming the DRIP queen! Thank you for being a great advisor. I appreciate all your guidance and advice you have given me these past five years. I am forever grateful for all the opportunities that I got to do as a graduate student. You have been my role model as a scientist and I hope that I have made you proud. I know it probably wasn't easy to have a persistent student who was constantly bugging you for something. Whether it was applying for a fellowship or a conference, you gave me your unconditional support and never said no when I asked for a letter of recommendation. Thank you for the time you have spent with me having intense scientific conversations. We have probably had a one hour scientific discussion every week for the past 5 years. That is over 260 hours of your time! Most advisors would not dedicate that time to their students and I am grateful that I learned so much from you during our conversations and reaffirmed why I love biology. I am thankful that we had the opportunity to attend a keystone conference together. There are many instances where I can recollect the kind and amazing mentor you are, but I will always remember that trip and all the amazing networking opportunities that stemmed from it because of you. Thank you for being my mentor. I hope that in the future I can be a supportive mentor like you. Someone who always looks out for their students and does their best to meet their needs. Thank you, Fred, for your continuous support during these past five years. Your guidance has helped me in many ways and I could not have imagined having a better advisor and mentor for my Ph.D studies.

ABSTRACT

Efficient co-transcriptional splicing is thought to suppress the formation of genome-destabilizing R-loops upon interaction between nascent RNA and the DNA template. In this work, it is demonstrated that inhibition of the SF3B splicing complex using Pladienolide B (PladB) in human K562 cells caused widespread intron retention and nearly 2,000 instances of R-loops gains. However, only minimal overlap existed between these events, suggesting that unspliced introns by themselves do not cause excessive R-loops. R-loop gains were instead driven by readthrough transcription at a subset of stress-response genes, defining a new class of aberrant "downstream of genes" (DoG) R-loops. Such DoG R-loops were temporally and spatially uncoupled from loci experiencing DNA damage. Unexpectedly, the predominant response to splicing inhibition was a global R-loop loss resulting from accumulation of promoter-proximal paused RNA polymerases and defective elongation associated with premature termination. Thus, SF3B1-targeted splicing inhibition triggered profound alterations in transcriptional dynamics, leading to unexpected disruptions in the global R-loop landscape.

CHAPTER I: INTRODUCTION

This chapter includes text that was published in DNA Repair (Amst) Volume 106, 103183, October 2021.

Castillo-Guzman D, Chédin F. Defining R-loop classes and their contributions to genome instability. doi:10.1016/j.dnarep.2021.103182. PMID: 34303066.

The original manuscritpt has been modified to rename the figures based on the format of this dissertation. My contributions to the original manuscript included writing of the introduction and editing of the entire manscript. I also contributed to the figure making of all the figures in the original manuscript. The figures were not included in this dissertation.

1.1 Overview

Gene expression in eukaryotes is a complex and tightly regulated multistep process. It involves transcription, mRNA maturation and splicing, mRNA export, translation, and degradation of mRNAs and proteins. Misregulation or alteration at any one of these steps can be detrimental to the cell, so understanding how cells maintain their gene expression programs is important for both basic biology and health. In addition, regulation of gene expression is the basis for cellular differentiation, development, morphogenesis, and the versatility and adaptability of any organism. At the foundation of gene expression programs is the DNA molecule, the source of all information necessary for life in every species. Maintaining stability of the genetic information is critical for the generation of proper RNAs from the DNA template by transcription. There are many sources of DNA Damage, exogenous and endogenous, that can give rise to DNA lesions or genome instability. Some examples of exogenous sources include ultraviolet light, X-ray irradiation, and toxins, while endogenous sources can arise from both normal and altered cellular metabolism (Lindahl, 1993; Lindahl and Nyberg, 1972). For example, transcription, a complex process that results in the expression of a gene, introduces significant force or torque on the DNA fiber resulting in superhelical stress (Forth et al., 2013; Ma et al., 2013). This stress or superhelicity favors transitions to alternative non-B DNA structures (Kouzine and Levens, 2007), such as the formation of R-loops (Chedin and Benham, 2020). R-loops are three-stranded non-B DNA structures that form during transcription. The nascent RNA reanneals to the template DNA strand, forming an RNA:DNA hybrid and causing the non-template DNA strand to loop out in a single-stranded state. Studies have linked R-loops to both positive and negative roles in the cell (Costantino and Koshland, 2015). For example, R-loops associate with hyperassible chromatin, thus may help maintain open, active chromatin states (Sanz et al., 2016). Given the impact genome instability has on cell function and organism health, including the development of cancers (Abbas et al., 2013; Negrini et al., 2010), understanding how normal cellular and metabolic activity, such as R-loop formation, can lead to genome instability is important.
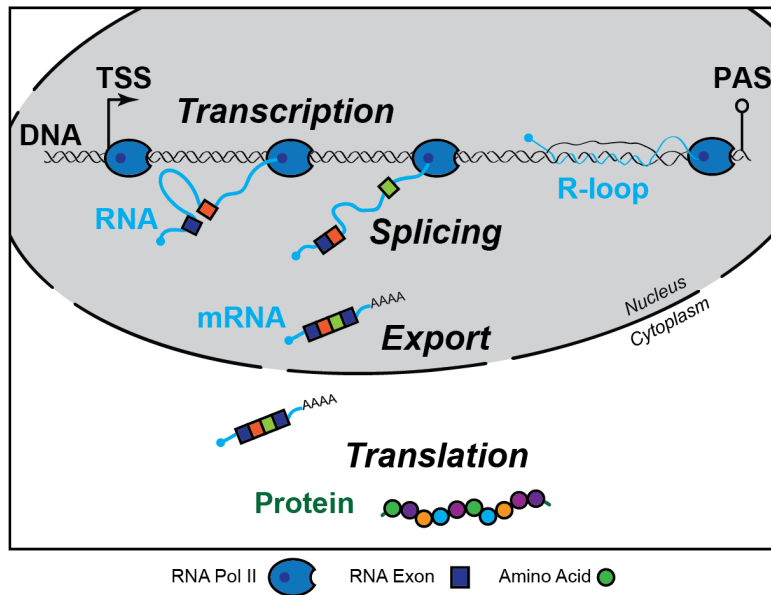
Figure 1.1 Schematic of gene expression pathway. Highlights transcription, splicing, mRNA export and translation.

Several studies have implicated R-loops as drivers of genome instability, but this model had never been directly tested at the genome-wide level. In my PhD work, the model that excessive co-transcriptional R-loops lead to genomic instability was tested and shown to be incorrect. In the following sections, key topics necessary for understanding R-loop biology and the questions addressed in dissertation will be introduced. A general overview of the gene expression pathway in higher eukaryotes will be provided, focusing on transcription, splicing, and mRNA export with pertinent background on R-loops and how they fit within the gene expression pathway (Figure 1.1). This will highlight how it is important to look at multiple parts of the gene expression pathway to understand the metabolism of R-loop structures. Current literature describing how R-loops may be sources of genome instability when mRNA biogenesis pathways are disrupted will be reviewed and used to identify the key gaps in knowledge my dissertation work aimed to address.

1.2 From DNA to a functional cytoplasmic mRNA

1.2.1 Transcription

During transcription, DNA is transcribed into RNA by a RNA polymerase complex. In eukaryotes, there are three types of RNA polymerases (RNA Pol). RNA Pol I is responsible for the majority of the transcriptional output as it transcribes tandem repeated array of genes encoding ribosomal RNAs (rRNAs) in nucleoli. RNA Pol III, which is located in the nucleoplasm, transcribes short non-translated RNAs necessary for ribosomal function such as transfer RNAs (tRNAs) and 5S rRNAs. Lastly, RNA Pol II, which is also located in the nucleoplasm, transcribes all the precursors of messenger RNAs (mRNAs) and a range of long and small non-coding RNA molecules. The intricate regulation of RNA Pol II transcription underlies cell differentiation and development via differential gene expression, contributing to the richness of eukaryotic form and function. From here on out, I will focus predominantly on RNA Pol II transcription and associated co-transcriptional processes such as splicing, and mRNA export.

Transcription consists of three main phases – initiation, elongation, and termination. At each of these steps, RNA Pol II is regulated and has extensive coordination with many essential accessory proteins. The RNA Pol II core catalytic complex consists of 12-subunits and is capable of unwinding DNA, synthesizing RNA, and rewinding DNA. RNA Pol II alone, however, is incapable of recognizing a promoter sequence and initiating transcription (Cramer, 2004). For this, it requires the participation of general transcription factors (GTFs), known as TFIIB, -D, -E, -F, and -H (Cramer, 2004; Hantsche and Cramer, 2016). In addition, other accessory factors are required for entering productive elongation and termination.

Initiation:

To initiate transcription GTFs must assemble with RNA Pol II near a core promoter. The core promoter encompasses 50 base pairs upstream and 50 bp downstream the transcription start site (TSS) (Hantsche and Cramer, 2016; Schilbach et al., 2021). At these core promoters the formation of the pre-initiation complex (PIC), which consists of RNA Pol II and six GTFs, occurs sequentially. There is

extensive evidence that the core promoter is first recognized by TFIID, followed by binding of TFIIA and TFIIB (Kraemer et al., 2001; Kuhlman et al., 1999; Nikolov and Burley, 1997). Then, RNA Pol II-TFIIF complex is recruited (Chen et al., 2007; Chen et al., 2010). Lastly binding of TFIIE and TFIIH follows (Maxon et al., 1994; Nikolov and Burley, 1997). TFIIH is 10-subunit protein complex that has a dual role in transcription initiation and in nucleotide excision DNA repair (NER) (Compe and Egly, 2016; Sainsbury et al., 2015). The XPB and XPD subunits both have helicase activity. XPB has $3´{\rightarrow}5´$ helicase activity, while XPD has $5´{\rightarrow}3´$ helicase activity (Rimel and Taatjes, 2018). This helicase activity is important to transition RNA Pol II from a closed to an open complex, where the TFIIH kinase DK7 can phosphorylate the RNA Pol II C-terminal domain (CTD) (Feaver et al., 1994; Lu et al., 1992; Meinhart et al., 2005). The CTD of RNA Pol II contains 52 heptad repeats with the consensus sequence $Y1\text{-}S2\text{-}P3\text{-}T4\text{-}S5\text{-}P6\text{-}S7$ (Harlen and Churchman, 2017). Throughout transcription the CTD is dynamically regulated by post-translational modifications, which can recruit or prevent the binding of key regulatory factors of RNA Pol II (Harlen and Churchman, 2017).

Promoter-proximal pausing and its regulation.

For years, it was thought that initiation was the major regulatory step during transcription, followed by monotonous and continuous transcription elongation. In metazoans, after PIC assembly and RNA Pol II has cleared the TSS, it transcribes 30-50 nucleotides downstream of the TSS before undergoing promoter-proximal pausing (Adelman and Lis, 2012). This pause is a key regulatory step under various conditions, as it primes RNA Pol II for entering productive elongation at the onset of a stimulus. Paused RNA Pol II was initially detected under heat shock conditions in heat-shock-responsive genes (Lis, 1998; Rougvie and Lis, 1988). Upon heat-shock, genes with primed RNA Pol II showed rapid release into productive elongation. The half-life of the paused RNA Pol II at different promoters can vary, a recent study reports half-lives ranging from ~2 to 30 minutes (Core and Adelman, 2019). Others have reported from several minutes to an hour or more (Jonkers et al., 2014; Krebs et al., 2017). In addition, stalled or paused RNA Pol II seems to block new transcription initiation and helps prevent new rounds of initiation (Shao and

Zeitlinger, 2017). The paused RNA Pol II is stabilized by the protein complexes negative elongation factor (NELF) and DRB sensitivity inducing factor (DSIF). NELF binds to the polymerase funnel to restrain RNA Pol II mobility that is require for pause release (Vos et al., 2018). The sequence downstream of the TSS may play a role, as core promoters of strongly paused genes often have elevated CG content (Hendrix et al., 2008) (see R-loop formation section below). Another model for how RNA Pol II pausing suggests that promoter-proximal pausing is an inherent property of transcription and occurs independent of the promoter sequence downstream of the TSS, which can still influence the stability of the interactions between DNA, nascent RNA, and paused RNA Pol II (Landick, 2006).

Many key factors are involved to transition the paused RNA Pol II into productive elongation. Release from promoter-proximal pausing involves phosphorylation of the RNA Pol II CTD at Ser 2 residues by the cyclin-dependent kinase 9 (CDK9) subunit of the positive transcription elongation factor b (P-TEFb). CDK9 forms a heterodimer with either Cyclin T1 or Cyclin T2 to form the positive transcription elongation factor-b complex, P-TEFb. Interestingly, the splicing factor SRSF2 promotes CDK9 activation through direct interaction with 7SK, a small ribonucleoprotein complex that binds to and inactivates P-TEFb, at exonic recognition sequences (Ji et al., 2013), suggesting that splicing dynamics are important for efficient transcription elongation. Other factors involved in enforcing the paused transcription complex are also phosphorylated by CDK9, including NELF and DSIF (Henriques et al., 2013; Lis et al., 2000), which results in the removal of NELF, while DSIF travels with RNA Pol II and transitions into an elongation factor.

The activity of P-TEFb is highly regulated and is part of at least three larger complexes. P-TEFb can interact with the super elongation complex (SEC) (Dahlberg et al., 2015). P-TEFb is also recruited to core promoters by the transcriptional cofactor BRD4, which is a member of the Bromodomain and Extraterminal (BET) protein family. BRD4 interacts with P-TEFb through its C-terminal domain (Dey et al., 2009; Schroder et al., 2012; Yang et al., 2008; Yang et al., 2005) and prevents P-TEFb from binding to the inhibitory ribonucleoprotein complex 7SK/HEXIM that sequesters P-TEFb in its inactive form (Jang et al., 2005; Schroder et al., 2012). BRD4 accumulates on hyper-acetylated and transcriptionally active TSSs and this accumulation functions as docking sites for P-TEFb. The acetyltransferase p300 and Pol II-

associated factor 1 (PAF1) have also been reported to be involved in RNA Pol II pause release (Boija et al., 2017; Imhof et al., 1997).

Elongation

The transition from initiation to a stable, elongating, RNA pol II complex occurs upon release of the RNA polymerase from the paused state. The elongation phase of the gene expression program is far from a simple, monotonous event and it has recently become clear that gene expression is also regulated at the level of elongation (Bentley and Groudine, 1986; Eick and Bornkamm, 1986). Elongation speed, for instance, is tightly regulated both between genes and within genes, as slower elongation correlates with higher splicing efficiencies (Jonkers et al., 2014; Kwak et al., 2013; Kwak and Lis, 2013). Reduced elongation can eventually lead to transcription pausing (Jonkers and Lis, 2015), which can either hinder or assist RNA processing or even lead to premature transcription termination at alternative gene ends.

At the rudimentary level, the elongation complex is composed of DNA, nascent RNA, and RNA Pol II. To describe the process of elongation requires delving into the mechanistic conformational changes RNA Pol II undergoes in transcribing DNA into RNA. During this process, RNA Pol II moves forward to produce RNA by a 'Brownian ratchet' translocation mechanism (Dangkulwanich et al., 2013; Nudler, 2012), where the polymerase oscillates between a pre-translocation state and a post-translocation state. In the pre-translation state, the active site of RNA Pol II contains base-paired RNA and DNA, while in the post-translocation state, the active site contain template DNA that is ready to be paired with a nucleoside triphosphate (NTP). In the post-translocation state (forward translocation), the incorporation of the NTP is thermodynamically favorable and generates a new RNA 3′-end causing the cycle to reset and repeat itself as RNA Pol II moves forward (Abbondanzieri et al., 2005; Bar-Nahum et al., 2005). Because of the 'Brownian ratchet', the moments of active transcription elongation are separated by transient pauses (Adelman et al., 2002). The RNA Pol II can undergo a reverse-translocation movement, where the active site of RNA Pol II and the 3′-end of the nascent RNA may become disengaged, in a process known as backtracking (Ishibashi et al., 2014; Nudler et al., 1997). This movement renders RNA Pol II inactive, but stable, providing a mechanism for many regulatory pauses and arrests. For example, when RNA Pol II

encounters unrepaired DNA lesions, backtracking allows repair enzymes to access the template (Mullenders, 2015). Arrested complexes can be rescued by TFIIS (Izban and Luse, 1992), which induces RNA cleavage by displacing RNA from the backtrack site within RNA Pol II (Cheung and Cramer, 2011) and stimulates the endonuclease activity of RNA Pol II so that the active site of RNA Pol II engages with the 3´-end of the nascent RNA.

In addition to the intrinsic pausing behavior of RNA Pol II, other factors can influence elongation in a positive or negative manner. For example, factors can favor the forward translocation state, help stabilize the paused state, or allow RNA Pol II to escape from backtracking (Ishibashi et al., 2014), all of which modify or influence the dynamics of RNA Pol II.

Transcription occurs in the context of a nucleosomal template and many factors that regulate RNA Pol II elongation function by modifying the chromatin landscape, including factors involved in nucleosome turnover and co-transcriptional modifications of histones and DNA. The nucleosome is the fundamental unit of chromatin, and wraps ~147 bp of DNA around its histone core and can hinder transcription elongation (Lai and Pugh, 2017). The histone core, or octamer, has one H3-H4 tetramer and two H2A-H2B dimers (Kornberg and Lorch, 1999). During transcription, one H2A-H2B dimer is removed from the nucleosome, resulting in the sub-nucleosomal particle called the hexasome (Kireeva et al., 2002). The facilitates chromatin transcription (FACT) complex travels with RNA Pol II and interacts with H2A-H2B dimers to regulate disassembly and reassembly of nucleosomes (Bondarenko et al., 2006; Hondele et al., 2013; Saunders et al., 2003). There are other factors that help mediate transcription elongation through nucleosomes, such as SPT6, chromodomain-helicase DNA binding protein 1 (CHD1) and imitation SWI (ISWI) (Petesch and Lis, 2012; Teves et al., 2014). Histone tails and globular domains can be modified by a wide range of post-translational modifications. Some of these modifications are specifically associated with distinct stages of transcription. Actively transcribed promoters contain high levels of H3K4me3 and acetylated H3 and H4 residues. Actively transcribed gene bodies are enriched for mono-ubiquitylated H2B, H3K36me3, H3K79me2 and H3K79me3 (Fuchs et al., 2014; Guenther et al., 2007; Vakoc et al., 2006). SETD2 is a histone methyltransferase that methylates K36 on H3 and associates with the phosphorylated

Pol II CTD. Deposition of H3K36me3 in gene bodies can affect RNA splicing and prevent cryptic transcription (McDaniel and Strahl, 2017; Venkatesh and Workman, 2015). For example, this histone mark can be recognized by the PWWP domain of DNA (cytosine-5)-methyltransferase 3B (DNMT3B), which methylates DNA in gene bodies to help prevent cryptic transcription (Neri et al., 2017). However the functions of the other modifications are not well understood.

Enhancer-promoter interactions are also essential for regulating transcription elongation. Enhancers are segments of DNA that have the ability to activate gene expression over large genomic distances (Serfling et al., 1985). The sequence composition of enhancers contain a high density of DNA motifs which can be recognized by activating transcription factors (TFs) that bind and active P-TEFb to promote transcription elongation (Yang et al., 2005; Zhang et al., 2012), thus enhancers can also influence promoter-proximal pausing (Liu et al., 2013; Zippo et al., 2009). Although the structural organization of enhancer-promoter interactions is not fully understood, the cohesin and Mediator complexes have critical roles in mediating these interactions (Kagey et al., 2010; Phillips-Cremins et al., 2013; Rollins et al., 2004). There is evidence for cohesin-dependent enhancer activity that modulates the role of the SEC in transcription elongation. Mediator can alleviate the stabilization that Gdown, which is a substoichiometric subunit for RNA Pol II, provides for RNA Pol II pausing and allows TFIIF to function (Cheng et al., 2012; Jishage et al., 2012). Mediator can also recruit the SEC and BRD4 to promoter regions and stimulate RNA Pol II pause release (Donner et al., 2010; Galbraith et al., 2013; Takahashi et al., 2011). Pause release and processive elongation are intertwined processes, and there are multitasking factors involved in both processes. Once RNA Pol II reaches the 3´ end of the gene, elongation RNA Pol II must transition into transcription termination. During this step the RNA is cleaved and polyadenylated.

Termination

At the end of a gene, RNA Pol II must be ejected or removed from chromatin so the complex can be recycled. This process is initiated by RNA cleavage and coordinated with polyadenylation of the transcript's 3´-end. Multiple factors mediate RNA cleavage and polyadenylation, including the cleavage

and polyadenylation specificity factor (CPSF), and the cleavage stimulation factor (CSTF). The activity of these factors is triggered upon recognition of the poly-adenylation signal (5´-AAUAAA-3´) in the nascent RNA transcript. There are two predominant models for RNA Pol II transcription termination, the torpedo model and the allosteric model. In the torpedo model, cleavage of the 3´-end of the RNA, releases the pre-mRNA from RNA Pol II, exposing the 3´-end of the transcript to the action of the poly(A) polymerase and other poly(A) binding proteins that guide the formation of the characteristic poly(A) tails of mRNAs. The cleavage event also generates a novel 5´-end for a distal transcript that is still being actively synthesized by RNA pol II. This novel, uncapped, 5´-end is accessible to the 5´→3´ exoribonuclease XRN2, which chases down RNA Pol II and ultimately displaces it from the DNA template (Fong et al., 2015). Under this model, the elongation speed of the RNA Pol II dictates the termination point, since XRN2 will have to travel longer before it reaches the RNA Pol II complex (Fong et al., 2015). In the allosteric model, a conformational change of RNA Pol II triggered by polyA signal recognition triggers termination and dislodges RNA Pol II from chromatin (Epshtein et al., 2007). Because a gene can possess multiple polyA sites, differential use of these sites can lead to the formation of distinct mRNA isoforms. Alternative polyadenylation (APA) is very common in eukaryotes and ~70% of mammalian protein-coding genes can undergo APA (Derti et al., 2012; Hoque et al., 2013), leading to transcriptome and proteome diversity. In yeast, slow RNA Pol II mutants switch to use of upstream polyA sites (Geisberg et al., 2020; Liu et al., 2017; Pinto et al., 2011; Yague-Sanz et al., 2020), while fast RNA Pol II mutants tended to favor downstream polyA site usage (Geisberg et al., 2020), suggesting that RNA Pol II speed influences recognition of the polyA site and triggers transcription termination.

These two models have been used for many years to explain how transcription termination occurs at protein-coding genes. Eaton *et al.* (2020), recently proposed a unified allosteric/torpedo mechanism for termination that incorporates features of both models (Eaton et al., 2020). Consistent with the torpedo model, the polyA site is recognized by CPSF30 and WDR33, which are components of CPSF. The nascent pre-mRNA is cleaved by the CPSF73 endonuclease, where now there exposed 5´ end of the nascent RNA can be captured by XRN2 (Eaton and West, 2020). In addition, RNA Pol II undergoes allosteric changes,

which are mediated by dephosphorylation of Spt5 subunit of DSIF by the protein phosphatase 1 (PP1) (Eaton et al., 2020). This dephosphorylation slows down elongation two-fold (Cortazar et al., 2019) and allows for XRN2 to catch up to RNA Pol II, resulting in transcription termination(Cortazar et al., 2019; Eaton et al., 2020).

1.2.2 Splicing

A human gene on average has eight introns. These introns need to be spliced out to form a functional mature mRNA. This requires *de novo* assembly of a spliceosome on each intron that is spliced out, thus multiple spliceosomes are required to remove introns from the pre-mRNA body (Sakharkar et al., 2005; Wahl et al., 2009). Nascent RNA is mostly spliced during transcription elongation (Brugiolo et al., 2013), and there is evidence that splicing can occur as soon as the intron exits from RNA Pol II (Alpert et al., 2017; Oesterreich et al., 2016). This emerging evidence that the spliceosome is physically close to RNA Pol II during transcription, implies that transcription and splicing occur on similar timescales and have the potential to influence one another. In higher eukaryotes there are larger, more complex pools of splicing factors, that enable variable splicing outcomes, including frequent alternative splicing. In addition, core transcription and processing machineries can be modulated to obtain different splicing outcomes (Herzel et al., 2017). This complex process centers around recognition of the sequence features within an intron. The sequence structure of an intron is conserved in eukaryotes: the GU and AG dinucleotides contained within the sequence of the splice sites define the 5´- and 3´-intron boundaries, respectively. Within each intron, a sequence located 18-40 nucleotides upstream of the 3´-splice site defines the essential branch point sequence (BPS) (Gould et al., 2016; Mercer et al., 2015; Qin et al., 2016; Taggart et al., 2017). Importantly, despite the conservation of splice sites, the exact boundaries of introns and their usage remain difficult to predict, thus most introns are annotated using empirical data, such as RNA sequencing (Lim and Burge, 2001).

At the chemical level, splicing involves a two-step transesterification reaction (Reed and Maniatis, 1985; Wahl et al., 2009). In the first step (Step I), the 2´-OH of the BPS adenosine carries out a nucleophilic

attack on the 5´-splice site guanosine, resulting in a 5´-exon with a 3´-OH and a branched intron lariat that is still attached to the 3´-exon. In the second step (Step II), the newly exposed 3´-OH of the 5´-exon attacks the first nucleotide downstream of the 3´-splice site guanosine, resulting in the ligation of both exons and the excision of the intron lariat. These reactions are catalyzed by core components of the spliceosome, which include five U-rich small nuclear ribonucleoproteins (snRNPs), which are named after their small nuclear RNA (snRNA). The five snRNPs are: U1, U2, U4, U5, and U6. The snRNPs along with other splicing accessory factors assemble the spliceosome in a stepwise manner and coordinate the two-step transesterification reaction to ensure proper secondary and tertiary RNA structures to define the spliceosome catalytic center and position the pre-mRNA for the splicing reaction. In the so-called spliceosome assembly model (Wahl et al., 2009; Will and Luhrmann, 2011), the first stages of spliceosome assembly involve the identification of the 5´- and 3´-ends of the intron, followed by spliceosome maturation, spliceosome activation, and splicing catalysis. The spliced RNA is then released, the spliceosome disassembled and recycled, and the intron lariat is debranched and degraded (Wahl et al., 2009; Will and Luhrmann, 2011). There are eight main spliceosome complexes that form during spliceosome assembly: E, A, B, B^{act}, B*, C, C*, and P (Herzel et al., 2017).
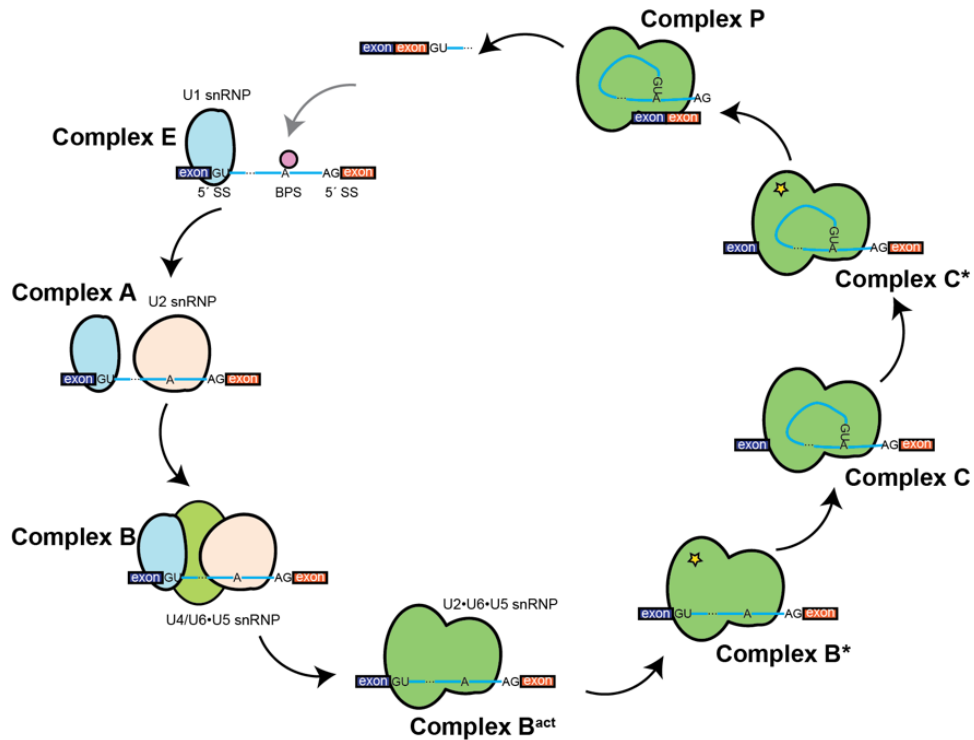
Figure 1.2 Spliceosome Aseembly during Transcription. General steps: U1 snRNP and BBP bind to form complex E. U2 snRNP binds to form complex A. U4/U6•U5 snRNP binds to form complex B. U1 and U4 leave resulting in formation of complex B^{act}. U2•U6•U5 remain bound and entire complex denoted by one green complex. U2•U6•U5 are activated to form complex B* to undergo the first splicing reaction on 5'SS, resulting in formation of complex C. This complex is then activated to form complex C*. Second splicing reaction occurs resulting in the ligation of two exons and removal of the intron to form complex P. Cycle can repeat itself. See below for details.

Assembly of the catalytically active spliceosome

Complex E forms when U1 snRNP binds to the 5´-splice site along with splicing factor 1 (SF1) and U2 auxiliary factor (U2AF), which help establish RNA-protein interactions with the BPS. The branch-point binding protein (BBP) also binds to the BPS (Baejen et al., 2014). Next BBP dissociates and U2 snRNA base pairs with the BPS to form complex A. This base pairing causes the BPS adenosine to bulge out and carry out the nucleophilic attack of 5´-splice site during Step I. Complex B forms when the tri-

snRNP (U4,U5, and U6) are recruited (Nguyen et al., 2015; Nguyen et al., 2016; Rigo et al., 2015; Wan et al., 2016b). Release of U1 and U4 results in spliceosome maturation and formation of the activated complex B^{act} (Fabrizio et al., 2009). At this point the spliceosome contains all the catalytic components but is still not active (Rauhut et al., 2016; Yan et al., 2016). At this stage, SF3b complexes of the U2 snRNP are in direct contact with the pre-mRNA upstream and downstream of the BPS and keep the BPS away from the catalytic center (Rauhut et al., 2016; Yan et al., 2016). To expose the BPS adenosine, the ATPase Prp2 destabilizes the interactions between SF3b and the BPS. This triggers the conversion of B^{act} into the active complex B* (Liu and Cheng, 2012; Rauhut et al., 2016). Upon activation, the BPS adenosine can now undergo nucleophilic attack on the 5′-splice site guanosine to complete step I. Next, complex C forms when the U5 snRNA base pairs with the 5′-exon to maintain the spliceosome active site, while U6 snRNA interacts with the branched 5′-splice site and the U2 snRNA immobilizes the intron lariat (Galej et al., 2016; Wan et al., 2016a). Activated complex C* forms through stimulation of the ATPase activity of Prp16 (Ohrt et al., 2013; Schwer, 2008). This activation results in the rearrangement of the catalytic site from step I to step II (Bertram et al., 2017; Fica et al., 2017; Yan et al., 2017). After step II is completed, it results in the formation of complex P, the post-spliceosome complex that contains the ligated 5′exon to 3′exon and the excised intron lariat. The last step requires the ATPase Prp43, which disassembles the intron lariat spliceosome into the intron lariat and the spliceosomal components (Fourmann et al., 2013). The components are recycled to repeat this process once the next intron exits out of RNA Pol II.

Transcriptional interactions that can influence splicing kinetics

Spliceosome assembly and catalysis often occurs in close proximity to the transcribing RNA Pol II, and this process is dependent on the proper recognition of the splice sites and BPS on the nascent RNA. Single molecule studies have shown that spliceosome assembly and catalysis take place when RNA Pol II has transcribed 26-129 nucleotides downstream of the 3′-splice site(Alpert et al., 2017; Oesterreich et al., 2016). Based on this data, only a portion of the RNA is available for the splicing machinery to recognize. In addition, transcription rates also determine the portion of the nascent RNA that is available for

spliceosome assembly. For example, faster transcription rates would produce longer stretches of nascent RNA sequence. Thus, RNA Pol II speed has to potential to regulate splicing outcomes and how the spliceosome assembles on nascent RNA. Studies have shown that events such as constitutive splicing, alternative splicing, back-splicing or alternative polyadenylation are regulated by RNA Pol II speed (Muniz et al., 2021). It is important to note that splicing dynamics themselves are variable, and splicing does not always follow the order of transcription. There is growing evidence that introns can remain on the nascent transcript and the order from which they are removed can vary. In addition, the kinetic process of splicing can span from minutes to hours (Drexler et al., 2020; Wan et al., 2021), If transcriptional dynamics can influence the stochastic nature of splicing is unclear.

Constitutive splicing occurs when two consecutive exons are joined after intron removal. Most introns are constitutively spliced. However, more than 95% of humans genes undergo alternative splicing (Adkins and Tyler, 2006), which contributes to the proteome diversity and can result in the formation of several mRNA splice variants from the same pre-mRNA. Both constitutive splicing and alternative splicing are sensitive to RNA Pol II speed. For example, the rates of elongation are faster over introns than over exons. Studies in yeast showed that slow RNA Pol II mutants had enhanced splicing, while fast mutants displayed reduced splicing efficiencies (Aslanzadeh et al., 2018; Braberg et al., 2013). Similar results were obtained from studies analyzing slow RNA Pol II mutants in *Drosophila* (Aslanzadeh et al., 2018; Braberg et al., 2013; Howe et al., 2003; Khodor et al., 2011). In plants, however, it was fast RNA Pol II mutants that showed enhanced constitutive splicing (Leng et al., 2020). In humans, both slow and fast RNA Pol II mutants inhibit constitutive splicing (Fong et al., 2014). These studies show that while the regulation of the rate of RNA Pol II progression is essential for proper regulation of constitutive splicing efficiency, the manner in which the rate influences splicing differs dramatically between species.

1.2.3 mRNP Processing and mRNA export

During transcription the RNA undergoes 5´-end capping, splicing to ensure removal of intron sequences, and 3´-end processing to create a polyadenylated end. Once a mature mRNA is produced it is

usually delivered to the cytoplasm, where it can be translated into protein. For the mRNA to be exported out of the nucleus, it has to form an export-competent messenger ribonucleoparticle (mRNP) (Moore and Proudfoot, 2009), which functions to protect the mRNA from degradation and to prepare it for export through a nuclear pore complex (NPC). mRNP formation is coupled to transcription and requires export factors. There are two main export receptors, NXF1 and CRM1, which interact with different sets of export adaptors. The NXF1 export pathway involves the recruitment of the TREX (transcription/export) complex to the mRNA during transcription, followed by recruitment of other adaptors, SR proteins, which ultimately results in the recruitment of NXF1 to the mRNP and delivery to the cytoplasm through the NPC transport channel.

The TREX (transcription/export) complex plays a key role in mRNP biogenesis and export (Aguilera, 2005a; Kohler and Hurt, 2007; Reed and Cheng, 2005; Reed and Hurt, 2002). It includes the THO complex and the helicase UAP56. The THO complex (a subcomplex of TREX) is recruited to chromatin during transcription and is needed for transcriptional elongation, 3´-end processing, nuclear RNA export and genome stability. This recruitment is facilitated through the association with the Ser2 phosphorylated CTD of RNA Pol II (MacKellar and Greenleaf, 2011; Meinel et al., 2013). In *Saccharomyces cerevisiae*, THO is composed of 5 subunits, while in higher eukaryotes it has 6 proteins: THOC1, THOC2, THOC3, THOC5, THOC6 and THOC7 (Masuda et al., 2005). THOC1 contains a death-like domain and the function of this domain is still unclear. Mammalian THOC5 exhibits a PEST like domain that contains three ATM kinase specific phosphorylation sites (Ramachandran et al., 2011). Human THO colocalizes with splicing factor in nuclear speckle domains *in vivo* (Masuda et al., 2005). In mammalian cells, THOC1 and THOC5 are mainly nuclear, though there is evidence that THOC5 can shuttle between the nucleus and cytoplasm (Katahira et al., 2009).

Transcription, splicing, and mRNA export are coordinated and essential to ensure mature mRNAs can be exported out of the nucleus into the cytoplasm. Once in the cytoplasm they can be used to generate proteins, stored, decayed, or even secreted. All of these steps have to be integrated and have to occur with proper kinetics for gene expression to be successful. There are mechanisms to help surveilleance the quality

of nuclear mRNA to prevent the export of potentically problematic transcripts. In broad terms this involves either retention of aberrant mRNAs or rapid degration of aberrant mRNAs. For example, inhibition of splicing can cause transcripts to remain chromatin bound (de Almeida et al., 2010; Martins et al., 2011). For the mispliced transcripts that are polyadenylated and released from the chromatin, they can accumulate in subnuclear domains called nuclear speckles (Girard et al., 2012; Martins et al., 2011), where they can potentially undergo post-transcriptional splicing and be release and exported (Girard et al., 2012). Abberants mRNA transcripts are degraded by the nuclear exosome, which has both 3´→5´ exonuclease and endonuclease activities (Chlebowski et al., 2013; Schneider and Tollervey, 2013). Another proposed role for the exosome is to reduce R-loop accumulation by preventing interactions between nascent RNA and exposed single stranded DNA (ssDNA) (Pefanis et al., 2015).

1.3 Transcription and R-loops as a source of genome instability

In recent years, growing evidence suggests that the down-regulation or inhibition of many factors involved in key steps in the gene expression pathway results in the global increase of R-loops. In addition, there is evidence that the global increase in R-loops results in genome instability. Thus, the focus of my dissertation work was to understand how the misregulation of R-loops contributes to genome instability when splicing is inhibited. In the following section I will go in more detail on background pertaining to R-loop metabolism, how we map these structures, and the evidence implicating them as sources of genome instability.

Genomes experience superhelical stresses during transcription. Transcription requires the participation of large macromolecular enzymes that as described above, translocate processively along the DNA (Cozzarelli et al., 2006). As RNA Pol II and accompanying accessory factors translocate along the DNA, they impose large amounts of torque on the fiber (Forth et al., 2013; Ma et al., 2013). Positive supercoils accumulate ahead of the advancing RNA polymerase and negative supercoils form behind the RNA polymerase. This is referred to as the "twin supercoiling domain model" (Liu and Wang, 1987; Ma and Wang, 2016).

DNA topoisomerases, which are a family of ubiquitous and conserved proteins (Lee, 2011; Pommier et al., 2016), can transiently cut, untwist, and religate DNA strands to relax superhelicity. Despite the action of DNA topoisomerase, transcription still results in high levels of torsional stress, especially for highly transcribed genes. Actively transcribed genes often experience transcription-associated DNA breaks (Kim and Jinks-Robertson, 2012) and recombination (TAR) (Gottipati and Helleday, 2009). Another consequence of torsional stress is that it favors the formation of non-B DNA structures such as R-loops, which can help store the negative superhelicity and return the DNA to an energetically favorable state (Chedin and Benham, 2020).

R-loops are three-stranded non-B DNA structures that form during transcription upon reannealing of the nascent RNA to the template DNA strand, forming an RNA:DNA hybrid and causing the non-template DNA strand to loop out in a single-stranded state. R-loops were first recognized to form at the replication origins of bacterial ColE1-type plasmids, where they serve to open the DNA double helix and the RNA strand can be processed upon Ribonuclease H digestion into a primer for leading strand replication (Itoh and Tomizawa, 1980; Masukata and Tomizawa, 1990). A similar mechanism was shown to mediate DNA replication initiation in bacteriophage T4 (Belanger and Kreuzer, 1998; Carles-Kinch and Kreuzer, 1997; Kreuzer and Brister, 2010) and in the mitochondrial genome (Lee and Clayton, 1996, 1998; Xu and Clayton, 1995). R-loops were then recognized to form in the chromosomes of mammalian B cells upon induction of transcription at specialized class switch regions (Yu et al., 2003). In this case, R-loop formation is associated with the formation of programmed double-stranded DNA breaks (DSBs) that are required to initiate immunoglobulin class switch recombination. Thus, from early on, it became apparent that transcription-mediated R-loop formation could play important physiological roles from *E. coli* to mammals, and that R-loops represent a novel type of *cis*-acting biological signal.

The study of *E. coli* RNase H mutant strains provided evidence that R-loops, if left to accumulate, could cause significant problems. *rnhA* mutants have the unique ability to replicate their genome independently of the chromosomal replication origin *oriC* and of the DnaA replication initiation protein (Kogoma and von Meyenburg, 1983; Ogawa et al., 1984). A similar ability was also observed for knockout

mutants of the *recG* gene, which encodes a helicase capable of resolving R-loops (Hong et al., 1995). This mode of replication termed constitutive stable DNA replication (cSDR) is strictly dependent on the recombinase activity of the RecA protein (Kogoma et al., 1985). It arises due to the formation of RecA-catalyzed R-loops that persist due to their reduced resolution in the absence of RNase H or RecG activity, and initiate DNA replication at alternative replication origins termed oriKs distributed along the chromosome (Drolet and Brochu, 2019; Kogoma, 1997). *rnhA*⁻ and *recG*⁻ *E. coli* mutants show sluggish growth, and increased genome instability, consistent with the induction of replication forks from oriKs causing global alterations of replication migration patterns (Maduike et al., 2014; Wimberly et al., 2013). A similar induction of alternative replication origins was observed at highly transcribed rDNA regions in yeast cells that accumulate R-loops due to deficiency in RNase H activity and DNA topoisomerase I (Stuckey et al., 2015). Therefore, it became clear early on that R-loop levels must be tightly controlled to avoid deleterious consequences on genome stability and that cells have evolved enzymes such as RNase H and helicases to promote R-loop resolution.

Over the last decade, our understanding of R-loops, including the mechanisms that control their formation and resolution, genomic distribution, and functional consequences has dramatically increased. R-loops represent a prevalent class of non-B DNA structures in all genomes including yeasts, plants, flies, and worms (Alecki et al., 2020; El Hage et al., 2014; Hartono et al., 2018; Wahba et al., 2016; Xu et al., 2020; Xu et al., 2017; Zeller et al., 2016). In mammalian genomes, R-loops collectively occur at tens of thousands of conserved genic loci (Chen et al., 2017; Crossley et al., 2019b; Ginno et al., 2012; Sanz et al., 2016; Wang et al., 2021b; Yan et al., 2019), highlighting the fact that R-loops are well-tolerated by cells under normal conditions. In addition, a variety of functional roles such as that described above for prokaryotic replication origins, have been assigned to R-loops, further suggesting that they play adaptive roles under physiological situations (see below).

At the same time, many studies have suggested that under pathological conditions, harmful R-loops arise from defective cellular processes and trigger DNA damage and genomic instability. Defects in co-transcriptional processes such as splicing, polyadenylation, and RNA binding protein (RBP) recruitment

leading to export competence have been particularly associated with harmful R-loops (Aguilera and Garcia-Muse, 2012; Chan et al., 2014; Kaneko et al., 2007; Li and Manley, 2005, 2006; Paulsen et al., 2009; Stirling et al., 2012). One of the key pieces of evidence supporting the idea of harmful R-loops is that cellular over-expression of Ribonuclease H1 (RNase H1), an enzyme with a clear biochemical ability to resolve RNA:DNA hybrids and R-loops (Cerritelli and Crouch, 2009), can at least partially suppress a variety of genome instability phenotypes (Huertas and Aguilera, 2003; Paulsen et al., 2009). Harmful R-loops, in turn, were proposed to affect genome stability by causing or exacerbating transcription-replication collisions (Hamperl et al., 2017; Hamperl and Cimprich, 2014, 2016; Lang et al., 2017), triggering replicative stress (Barroso et al., 2019; Crossley et al., 2019a; Herold et al., 2019; Landsverk et al., 2019; Morales et al., 2016), or inducing nuclease-mediated DNA breakage (Sollier and Cimprich, 2015; Sollier et al., 2014). We note, however, that the association between harmful R-loops and genome instability relied in many instances on observations of excessive R-loop levels by S9.6 immunofluorescence microscopy. Recent evidence, however, suggests that these observations may need to be revisited given the likelihood of significant confounding artefacts in S9.6 imaging studies (Smolka et al., 2021a). At the genomic level, harmful R-loops remain poorly characterized. Similarly, the spatiotemporal relationship between harmful R-loop formation, DNA damage initiation and their suppression by RNase H1 expression has for the most part never been directly assessed. Thus, significant questions remain surrounding the identities of harmful R-loops and their mechanism of action. In addition to increase in DNA damage, other mechanisms may be at play under and how defects in RNA binding processes impact nuclear homeostatsis. Here, we focus on reviewing recent R-loop mapping efforts in mammalian cells. We suggest that these studies can be most easily reconciled considering the existence of distinct R-loop classes, each with unique characteristics. We further propose that events of genome instability may be connected to specific R-loop sub-types.

1.3.1 R-loop mapping efforts suggest the existence of two classes of R-loops

Two main types of R-loop mapping methodologies have been developed to provide population-average views of genomic R-loop distributions. These strategies rely either on the S9.6 anti RNA:DNA

hybrid monoclonal antibody (Boguslawski et al., 1986) or on catalytically inactive variants of Ribonuclease H1 (dRNase H1) that are still binding-competent due to the RNase H1 hybrid-binding domain (Chen et al., 2017). Despite significant concerns about the use of S9.6 in imaging applications, it permits accurate R-loop mapping in genomics applications after DNA:RNA ImmunoPrecipitation (DRIP) (Smolka et al., 2021a). Several variations of the initial DRIP-seq method (Ginno et al., 2012) with various degrees of resolution and strand-specificity have been published (Crossley et al., 2019b; Sanz et al., 2016; Smolka et al., 2021a; Xu et al., 2017) (Chapter II focuses on the protocol used for mapping R-loop structures) and generally produce highly congruent maps in human cells (Chedin et al., 2021a). Importantly, S9.6-based methods require initial steps of DNA extraction and fragmentation which allow the pre-treatment of extracted nucleic acids by exogenous RNase H. As expected, DRIP-seq maps are highly sensitive to RNase H pre-treatment, providing an essential specificity control. In addition, DRIP-based maps have been independently validated using approaches based on non-denaturing sodium bisulfite in an S9.6-independent manner (Malig et al., 2020a). Nonetheless, S9.6-based mapping methodologies can be considered as mapping R-loops *ex vivo* since they require initial nucleic acid extraction from cells. By contrast, dRNase H1-based approaches rely on mapping R-loops either through mapping the binding sites of dRNase H1 expressed *in vivo* (such as in RNase H1 ChIP, or R-ChIP (Chen et al., 2017)) or by liberating R-loops from native chromatin *via* methodologies derived from CUT&RUN (short for "Cleavge Under Targets & Release using Nuclease") and CUT&TAG (short for "Cleavage Under Targets & Tagmentation) (Wang et al., 2021b; Yan et al., 2019). One key advantage of such methods is that R-loops are profiled under native conditions without the need to extract nucleic acids or chromatin from cells prior to mapping.

Distribution of R-loops from native mapping methodologies.

Major differences have emerged between dRNase H1- and S9.6-based R-loop maps. dRNase H1-based maps consistently identify R-loops over GC-rich and GC-skewed promoter-proximal pause regions of numerous transcribed genes. Of the twelve thousand or so R-ChIP-seq peaks recovered, nearly 60% mapped to promoter-proximal regions, significantly higher than observed in gene bodies (17%) or over

gene terminal regions (6%) (Chen et al., 2017). In addition to genic R-loops, dRNase H1-based methods have consistently detected the presence of several thousand intergenic R-loops mapping to active enhancer regions (Chen et al., 2017; Wang et al., 2021b; Wulfridge and Sarma, 2021; Yan et al., 2019). tRNA genes also found among the strongest hotspots for dRNase H1 binding by R-ChIP (Chen et al., 2017). This observation holds also true in yeast (El Hage et al., 2014; Hartono et al., 2018; Legros et al., 2014), which suggests that high R-loop loads may be associated with these short genes. Median R-loop peak sizes reported by R-ChIP were relatively short, around 200-300 bp (Chen et al., 2017; Wang et al., 2021b). Importantly, dRNase H1 recruitment to transcription start sites (TSSs) was dynamically correlated with transcriptional pausing, suggesting a mechanistic connection between pausing and R-loop formation (Chen et al., 2017).

Distribution of R-loops ex vivo.

S9.6-based maps, on the other hand, show that R-loops are predominantly distributed along transcribed genic regions and correlate with both gene expression levels and gene length (Sanz et al., 2016). Tens of thousands of conserved peaks of R-loop formation have consistently been recovered from a variety of human cell lines and from multiple studies (Chedin et al., 2021a; Sanz et al., 2016). About half of these peaks map to transcribed gene bodies, with hotspots observed downstream of GC-skewed CpG island promoters (13%) and gene ends (19%) (Sanz et al., 2016). By contrast with native R-loops that are confined to a short region immediately downstream of the transcription start site (TSS), DRIP approaches are only able to detect R-loops about 1-1.5 kb downstream of the TSS. While *ex vivo* promoter R-loops clearly associate with GC skew, we note that the degree of this sequence characteristic progressively decreases past the exon1 / intron 1 junction (Hartono et al., 2015), suggesting that other properties in addition to the thermodynamic stability of RNA:DNA hybrids may be facilitating R-loop formation downstream of the TSS. It is possible that as the RNA Pol II enters productive elongation, it introduces negative supercoiling on the DNA template, driving the formation of R-loops to relieve the associated topological stress (Stolz et al., 2019a). Interestingly, gene body and terminal *ex vivo* R-loops correlate with variable levels of GC skew,

suggesting that such an interplay between DNA sequence features and topological considerations may be at play for many loci (Chedin and Benham, 2020; Sanz et al., 2016). In contrast to native R-loops, *ex vivo* R-loops show little signal over tRNA genes and are not readily detected over intergenic enhancers. Similarly, *ex vivo* R-loops often define much larger peaks, with median lengths of 1.5 kilobases (Sanz et al., 2016). Single-molecule R-loop footprinting analysis revealed that such large peaks are caused by the clustering of smaller individual R-loops over larger R-loop zones (Malig et al., 2020a).

Reconciling R-loop classes: paused- versus elongation-associated R-loops.

The variation between S9.6- and dRNase H1-based methods could be explained by differences in specificities between RNase H1 and S9.6 and/or by the possibility that RNase H1 was targeted to R-loops found at paused promoters. To clarify the differences between dRNaseH1- and S9.6-mapping methods, a recent study profiled R-loops using the CUT&TAG technology, taking advantage of the N-terminal hybrid binding domain (HBD) of RNase H1 and of the S9.6 antibody as R-loop sensors that were fused to GST- and His6-tagged moieties (Wang et al., 2021b). Each sensor protein was then used for both native and *ex vivo* R-loop profiling. The HBD sensor protein, when used *ex vivo*, generated maps similar to those obtained using the S9.6 sensor, recapitulating previous high-resolution strand-specific profiling results using the DRIPc-seq methodology (Sanz and Chedin, 2019a; Sanz et al., 2016). This suggests that the RNase H1 HBD and S9.6 can recognize the same subset of R-loops. Strikingly, when used in CUT&TAG approaches for native and fragmentation-free R-loop mapping, both sensor proteins generated results consistent with other dRNase H1-based R-loop profiles. This establishes that the primary difference between S9.6-based and dRNase H1-based R-loop mapping derives from the application of these reagents to mapping R-loops in a native context *versus ex vivo*. Methods that capture native R-loops like MapR, R-ChIP, and R-loop CUT&TAG better reflect R-loops formed near paused promoter regions, while methods that capture R-loops *ex vivo* like DRIP-seq and its derivatives identify R-loops that form through gene body regions and therefore associate with transcription elongation. The mechanistic connections between these two R-loop types and transcriptional pausing *versus* transcription elongation are well-reflected in their response to

- 23 -

drugs such as 5, 6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB), which inhibits RNA Pol II by causing premature transcription termination, that enforce heightened promoter pausing by blocking the release of RNA Pol II into elongation. DRB treatment caused increased dRNase H1 recruitment to promoter regions, reflecting increased pausing-associated R-loops. Conversely, washes following DRB treatment, caused the reduction of dRNase H1 binding and promoter-associated R-loops, as expected from the release of previously paused RNA Pol II complexes into elongation (Chen et al., 2017). In sharp contrast, DRB treatment caused a rapid reduction of R-loops 1-2 kilobases downstream of promoters as profiled by DRIP-qPCR (Sanz et al., 2016), reflecting a rapid decrease in elongation complexes. Prolonged DRB treatment progressively suppressed all instances of R-loop formation along gene bodies (Crossley et al., 2019a; Sanz et al., 2016). As expected, washes following DRB treatment caused a rapid return of R-loops as measured by DRIP-qPCR, consistent with the resumption of elongation (Sanz et al., 2016). Thus, emerging data suggest that there are two distinct classes of R-loops that: (i) associate with two distinct states of the transcription cycle; and (ii) are best profiled through different approaches.

1.3.2 Contrasting properties of R-loop classes

Based on the R-loop mapping data, the proposed two R-loop classes possess distinct properties that may account for their differential ability to be detected. We note that RNA Pol I-driven and RNA Pol III-driven R-loops, both of which likely correspond to important R-loop classes, are not being discussed here. Similarly, the formation of RNA:DNA hybrids or R-loops at sites of DNA double-stranded breaks (Cohen et al., 2018; D'Alessandro et al., 2018; Ohle et al., 2016) is not considered here.

Length and stability.

Promoter-associated R-loops (referred to here as Class I) are expected to be small, reaching 60 bp at most given the lengths of RNA transcripts at promoter-proximal pause sites (Adelman and Lis, 2012). As suggested (Chedin et al., 2021a), the short lengths of such R-loops may result in lower stability during genome fragmentation in DRIP-based approaches. It is also possible that such small R-loops owe their

stability *in situ* to the presence of large protein complexes nearby, including the paused RNA Pol II machinery and associated pausing and pause-regulating factors. If so, deproteinization during *ex vivo* DNA extraction may further destabilize them. This, together with minimal size thresholds (>100 bp) enforced during DRIP library construction steps, may account for significant recovery losses over these regions in *ex vivo* approaches. Non-denaturing bisulfite-based approaches are similarly challenged in identifying Class I R-loops due to their short size and paucity of cytosines on the displaced strand (Chedin et al., 2021a; Malig et al., 2020a). It is possible that Class I R-loops can only be captured under native conditions. The difficulties associated with recovering and detecting Class I R-loops in *ex vivo* approaches may in fact have allowed the detection of more stable, but less abundant, elongation-associated (Class II) R-loops (see below). We note that paused RNA polymerases are often backtracked (Noe Gonzalez et al., 2021; Sheridan et al., 2019), and it was recently proposed that small "anterior R-loops" may form ahead of backtracked RNA polymerases (Zatreanu et al., 2019). The exact molecular features of promoter-associated Class I R-loops therefore remain to be clarified.

In contrast to Class I R-loops, single-molecule R-loop footprinting approaches revealed that Class II elongation-associated R-loops show median lengths of about 300 base-pairs and can extend to kilobase-length structures (Malig et al., 2020a). Thus, the two R-loop classes show nearly an order of magnitude difference in length. The large sizes of Class II R-loops may account for their relative stability to DNA extraction and fragmentation, allowing *ex vivo* profiling. As noted previously, however, it is likely that some Class II R-loops are unstable in the face of DNA fragmentation, especially when negative DNA supercoiling played a prominent role in driving their formation (Chedin and Benham, 2020; Stolz et al., 2019a).

Frequency of formation.

While Class I and Class II R-loops show clear differences in length and in their association with paused *versus* elongating RNA Pol II , much less is known regarding the frequency at which they form. The average yields for elongation-associated R-loops, measured by DRIP-qPCR as a percentage of input,

range from 1-10% at positive loci (Sanz et al., 2016). Yields from RchIP-qPCR are notably lower (Chen et al., 2017). We suspect, however, that this may not reflect the true frequency distribution of Class I and Class II R-loops. ChIP experiments, which involve crosslinking and harsh sonication prior to immunoprecipitation may be limited in their ability to efficiently recover Class I R-loops. This may be further compounded if only a portion of Class I R-loops are RNase H1-bound at any given time. Issues of epitope accessibility may further complicate recovery given the presence of large macromolecular complexes over paused promoter regions (Core and Adelman, 2019). While future experiments will be necessary to accurately quantify the relative amounts of Class I and Class II R-loops, we suggest that Class I R-loops formed over paused promoters are much more abundant than Class II R-loops are at any given position. If correct, this proposal suggests that "native" approaches are limited in their ability to recover Class II R-loops simply because the bulk of R-loops in a cell correspond to Class I R-loops formed at promoters. This proposal follows the well-accepted notion that the highest RNA Pol II density measured by ChIP-seq approaches, and the highest transcriptional activity measured by profiling nascent transcription, are primarily found over paused promoters compared to transcribed gene bodies (Henriques et al., 2013; Rahl et al., 2010; Wissink et al., 2019). Thus, the proposed high frequency of Class I R-loop formation may simply reflect the prevalence of promoter-proximal paused RNA Pol II complexes. Importantly, pause sites, particularly over CpG island promoters associate with very high, R-loop-favorable, GC skew levels (Chen et al., 2017; Hartono et al., 2015). In addition, the presence of a nearby free 5′-end may further facilitate R-loop initiation during promoter pausing (Chen et al., 2017; Roy et al., 2010). Finally, the fact that RNA Pol II machinery itself is paused may provide a long kinetic window for an R-loop to arise. Overall, we propose that Class I R-loops dominate the R-loop landscape by virtue of their association with abundant paused RNA Pol II complex, the presence of favorable sequence characteristics, and the availability of a free 5′-end.

Class II R-loops have an estimated half-life of about 10 minutes (Crossley et al., 2019b; Sanz et al., 2016); by contrast, the half-lifes of Class I R-loops are not known. It is reasonable to propose, however, that Class I R-loops may show a half-life similar to that of paused RNA Pol II complexes. Measurements of RNA Pol II pausing indicate that pause duration has a median value of 7 minutes (Jonkers et al., 2014), but that there exists considerable variation from 2 to 30 minutes depending on the promoter being considered (Core and Adelman, 2019). Enhancers may display an even shorter pause duration (Henriques et al., 2018). Thus, in all cases R-loop formation is a dynamic process but the half-lives, and potentially the types of enzymatic activities associated with R-loop resolution may vary between Class I and Class II R-loops.

1.3.3 Functional consequences of Class I and Class II R-loops

Class II R-loops have been associated with several important functions under normal conditions in mammalian cells (Chedin, 2016). Whether they occur in promoter-distal regions, gene bodies, or terminal genic regions, Class II R-loops correspond to regions of increased RNA Pol II density (Sanz et al., 2016). This suggests that they help to slow or stall the transcription machinery, as observed *in vitro* (Belotserkovskii et al., 2017; Belotserkovskii et al., 2018). Towards the beginning of genes, where Class II R-loops are prominent, slower elongation is expected to favor the recruitment of chromatin modifying enzymes to the C-terminal domain of RNA Pol II. This, in turn, may account for the increased deposition of several transcription-coupled histone modifications such as histone H3 lysine 36 trimethylation observed for R-loop-positive genes compared to expression-matched, but R-loop-negative, genes (Sanz et al., 2016). At the end of genes, which also correspond to Class II R-loop hotspots, R-loop-positive regions show dramatically elevated RNA Pol II stalling compared to expression-matched R-loop-negative terminal regions. Stalling, in turn, associates with efficient transcription termination, which is a property preferentially observed for genes with close neighbors (Sanz et al., 2016). Mechanistically, slower transcription elongation downstream of the polyadenylation site may tip the kinetic competition between

the XRN2 ribonuclease and RNA Pol II in favor of XRN2-mediated transcript degradation (Saldi et al., 2018). Class II R-loop formation also generally correlates with regions of increased chromatin accessibility, consistent with the notion that rigid A-like form RNA:DNA hybrids do not wrap around nucleosomes (Dunn and Griffith, 1980). Finally, R-loops were proposed to absorb large amounts of negative superhelicity, contributing to the transient relaxation of topological stresses in the genome (Chedin and Benham, 2020; Stolz et al., 2019a). Thus, under normal conditions, Class II R-loops have been assigned roles in chromatin patterning, transcription regulation, and topological management. Class I R-loops have been associated with chromatin features typical of highly active promoters over the promoter-proximal pause sites, such as high GC skew and G quadruplex motifs, high levels of H3K4 trimethylation and histone acetylation, and high RNA polymerase II occupancy (Chen et al., 2017). Interestingly, similar enrichments were observed around the TSSs of Class II R-loop-forming genes (Sanz et al., 2016), consistent with the notion that genes undergoing R-loop formation during transcription elongation also correspond to genes that accumulate Class I R-loops over their promoter-proximal pause regions. Indeed, Class II R-loop formation downstream of promoters was shown to be significantly associated with RNA Pol II pausing (Zhang et al., 2017).

1.3.4 Connections between different RNA binding factors and R-loops

While the physio-chemical bases for R-loop formation are becoming clearer, much remains to be learned about the cellular function and regulation of these structures. Studies from numerous labs have suggested that R-loops have both physiological and pathological outcomes (Costantino and Koshland, 2015; Skourti-Stathaki and Proudfoot, 2014). For instance, R-loops have been linked to efficient transcription termination (Sanz et al., 2016; Skourti-Stathaki et al., 2011), chromatin patterning (Chedin, 2016; Chen et al., 2015; Sanz et al., 2016), and the relaxation of negative DNA superhelicity (Stolz et al., 2019b), but R-loops have also been linked to genomic instability (Crossley et al., 2019a; Hamperl and Cimprich, 2014). What distinguishes "good" from "bad" R-loops remains mostly unclear, in part because the pathways regulating R-loop formation and resolution are incompletely understood.

Despite our growing understanding of where and when these structures form, many questions remain as to how these structures are regulated in the cell. One prominent model in the field is that RNA binding factors, such as those involved in splicing or mRNA export, have a role in suppressing the formation of R-loops genome-wide. In this model RNA binding factors bind to the RNA and sequester the RNA from interacting with the DNA to from an R-loop. If this model is correct, removal of RNA binding factors would cause an increase in R-loops genome-wide. The first evidence that an RNA binding factor may have a role in suppressing R-loops comes from the seminal study conducted by Li and Manley (2004); which showed that disruption of RNA splicing factors can result in R-loop-mediated genomic instability in human cells. This led to the hypothesis that RNA splicing factors, by sequestering the transcript away from the DNA template and reducing the homology between RNA and DNA, may play a role in preventing R-loops, thereby alleviating DNA replication stress, mutagenesis, and DNA breakage. Depletion of THOC1 and THOC5 led to an increase in double strand breaks and an increase in γH2AX (Cai et al., 2020; Dominguez-Sanchez et al., 2011; Katahira et al., 2009; Saran et al., 2016; Tran et al., 2016). Many of these studies implicate R-loops as a source contributing to DNA damage (Dominguez-Sanchez et al., 2011; Huertas and Aguilera, 2003; Luna et al., 2019). Overexpression of RNase H1 resulted in the decrease in γH2AX(Dominguez-Sanchez et al., 2011), but in some cases did not resolve the increase in γH2AX signal observed upon depletion (Cai et al., 2020). It should also be noted that in these studies, R-loops were detected using S9.6 immunofluorescence imaging and the genomic profiles of R-loops of THO depleted cells have never been characterized. While splicing factor mutations, elevated R-loop levels, and genomic instability have been implicated in a range of human disorders, the mechanistic relationships linking these processes remain largely to be elucidated.

1.3.5 Links between R-loops, Splicing, and Genome Instability

Mapping studies clearly indicate that R-loops often occur in introns (Sanz et al., 2016), indicating that the nascent RNA transcript engages with the template DNA prior to splicing. Evidence from yeast

indicates that introns may protect against R-loops owing to the fact that they are bound by RNA-binding factors involved in splicing (Bonnet et al., 2017). This led to a model whereby R-loop formation is reduced if the nascent transcript is engaged by RNA binding proteins, including the spliceosome, and the RNA export machinery (Santos-Pereira and Aguilera, 2015). Splicing, by removing introns, further reduces the homology between transcript and DNA template, thereby further reducing R-loop formation.

Interestingly, multiple studies (Gomez-Gonzalez et al., 2011; Paulsen et al., 2009; Santos-Pereira and Aguilera, 2015; Sollier and Cimprich, 2015) have shown that down-regulation of a large number of genes involved in RNA processing, including splicing factors (Aronica et al., 2016; Bhatia et al., 2014; Li and Manley, 2005; Onyango et al., 2017; Tresini et al., 2015; Wan et al., 2015), cause double-stranded DNA breaks and genomic instability through a mechanism thought to involve elevated R-loop levels. Overexpression of Ribonuclease H1 (RNase H1), an enzyme that degrades RNA when paired to DNA and resolves R-loops *in vitro* (Cerritelli and Crouch, 2009), suppresses the phenotypes linked to splicing factor knockdowns (Paulsen et al., 2009). Despite this, the genomic profiles of so-called "aberrant" R-loops have never been reported. Similarly, it has never been directly shown that R-loops cause or even positionally associate with unspliced RNA segments. Thus, the proposed role of splicing factors in regulating R-loop metabolism is mostly unclear.

1.4 SF3B1: a medically-relevant splicing factor target to analyze splicing / R-loops relationships

SF3B1 is an important component of the U2 snRNP particle that is frequently mutated in myelodysplastic syndromes (MDS) (Yoshida et al., 2011). Importantly, knockdown of several SF3b complex subunits led to the accumulation of the DNA damage marker gamma H2AX, linking U2 snRNP disruption to genomic instability (Paulsen et al., 2009). Similarly to what was originally observed by Li and Manley for the SRSF1 splicing factor (Li and Manley, 2005), over-expression of RNase H1 partially rescued the effect, suggesting "aberrant" R-loops underlie the instability phenotype (Paulsen et al., 2009). SF3B1 is therefore a highly relevant target to examine the mechanistic relationships between splicing dysfunction and R-loop metabolism. In addition, Pladienolide B (PladB) treatment offers a simple and

specific tool to inhibit SF3B1 pharmacologically and to determine the downstream consequences of U2 snRNP perturbation. PladB treatment specifically impairs SF3B1 function but allows the SF3B complex to maintain its physical interactions within the U2 snRNP particle (Kotake et al., 2007; Mizui et al., 2004; Yokoi et al., 2011). Specifically, U1 and U2 snRNP are able to bind to form an early version of complex A, but are unable to proceed with complex B formation (Figure 1.3) (Cretu et al., 2018; Effenberger et al., 2017; Finci et al., 2018; Jacobs et al., 2004). Thus, any change observed is due to the consequence of splicing defects alone. In addition, since PladB functions rapidly, it offers the possibility of measuring splicing, expression, and R-loop patterns in a dynamic fashion post treatment.
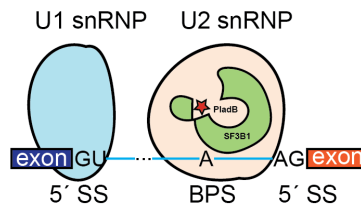
Figure 1.3 PladB mediated inhibition of SF3B1. PladB prevents the confirmational change of SF3B1, preventing the complete formation of complex A and transistion into complex B.

1.5 Dissertation project goals

The goal of my dissertation project was to determine the temporal and positional relationships between transcription, splicing and R-loop formation to clarify their mechanistic interconnections. To do this, I explored the mechanistic links between RNA splicing and R-loop formation using high throughput genomic methods and focused on the splicing factor SF3B1. It has been reported that inhibition of SF3B1 via PladB caused an increase in R-loops detected by S9.6 immunofluorescence (Chakraborty et al., 2018; Wan et al., 2015). Although there is evidence that PladB mediated splicing inhibition can lead to genomic instability, where the increase of R-loops is occurring is unclear. Another overlooked component is how inhibition of splicing affects transcription dynamics. Addressing how these processes globally respond to acute splicing inhibition is important to determine the mechanism(s) leading to genomic instability.

CHAPTER II: Mapping R-loops and RNA:DNA hybrids with S9.6-based immunoprecipitation methods

This chapter has been modified from the manuscript that was published under the same title in *J. Vis. Exp.* Volume 174. July 2021.

Sanz, L. A., Castillo-Guzman, D., Chédin, F. Mapping R-Loops and RNA:DNA Hybrids with S9.6-Based Immunoprecipitation Methods. *J. Vis. Exp.* (174), e62455, doi:10.3791/62455 (2021).

This manuscript has been modified to match the formatting of this dissertation. Figures and sections have been renamed. My contributions to the original manuscript include editing the text of the manuscript and Figure 2.2, outlining the script included in the video component of the manuscript.

2.1 Abstract

R-loops constitute a prevalent class of transcription-driven non-B DNA structures that occur in all genomes depending of both DNA sequence and topological favorability. In recent years, R-loops have been implicated in a variety of adaptive and maladaptive roles and have been linked to genomic instability in the context of human disorders. As a consequence, the accurate mapping of these structures in genomes is of high interest to many investigators. Here, we describe DRIP-seq (DNA:RNA Immunoprecipitation followed by high throughput sequencing), a robust and reproducible technique that permits accurate and semi-quantitative mapping of R-loops. We also describe a recent iteration of the method in which fragmentation is accomplished using sonication (sDRIP-seq), which allows strand-specific and high-resolution mapping of R-loops. sDRIP-seq thus addresses some of the common limitations of the DRIP-seq method in terms of resolution and strandedness, making it a method of choice for R-loop mapping.

2.2 Introduction

R-loops are three-stranded nucleic acid structures composed of an RNA:DNA hybrid and a single-stranded DNA loop. These structures form primarily during transcription upon hybridization of the nascent RNA transcript to the template DNA strand. Biochemical reconstitution (Daniels and Lieber, 1995; Ginno et al., 2012; Reaban et al., 1994; Yu et al., 2003) and mathematical modeling (Stolz et al., 2019a), in combination with other biophysical measurements (Carrasco-Salas et al., 2019; Duquette et al., 2004), have established that R-loops are more likely to occur over regions that exhibit specific favorable characteristics. For instance, regions that display strand asymmetry in the distribution of guanines (G) and cytosines (C) such that the RNA is G-rich, a property called positive GC skew, are favored to form R-loops when transcribed owing to the higher thermodynamic stability of the DNA:RNA hybrid compared to the DNA duplex (Huppert, 2008). Regions that have evolved positive GC skew, such as the early portions of many eukaryotic genes (Ginno et al., 2012; Green et al., 2003; Hartono et al., 2015; Polak and Arndt, 2008), are prone to forming R-loops *in vitro* and *in vivo* (Ginno et al., 2012; Malig et al., 2020b; Yu et al., 2003).

Negative DNA superhelical stress also greatly favors structure formation (Drolet et al., 2003; Masse et al., 1997) because R-loops efficiently absorb such topological stresses and return the surrounding DNA fiber to a favorable relaxed state (Chedin and Benham, 2020; Stolz et al., 2019a).

Historically, R-loop structures were considered to result from rare, accidental entanglements of RNA with DNA during transcription. However, the development of DNA:RNA immunoprecipitation (DRIP) coupled to high-throughput DNA sequencing (DRIP-seq) allowed the first genome-wide mapping of R-loops and revealed that those structures are far more prevalent than expected in human cells (Ginno et al., 2013; Ginno et al., 2012). R-loops occur over tens of thousands of conserved, transcribed, genic hotspots in mammalian genomes, with a predilection for GC-skewed CpG islands overlapping the first intron of genes and the terminal regions of numerous genes (Sanz et al., 2016). Overall, R-loops collectively occupy 3-5% of the genome in human cells, consistent with measurements in other organisms including yeasts, plants, fly and mouse (Alecki et al., 2020; El Hage et al., 2014; Hartono et al., 2018; Wahba et al., 2016; Xu et al., 2017).

Analysis of R-loop forming hotspots in human cells revealed that such regions associate with specific chromatin signatures (Chedin, 2016). R-loops in general, are found over regions with lower nucleosome occupancy and higher RNA polymerase density. At promoters, R-loops associate with increased recruitment of two co-transcriptionally deposited histone modifications, H3K4me1 and H3K36me3 (Sanz et al., 2016). At gene termini, R-loops associate with closely arranged genes that undergo efficient transcription termination (Sanz et al., 2016), consistent with prior observations (Skourti-Stathaki et al., 2011). R-loops were also shown to participate in the initiation of DNA replication at the replication origins of bacteriophage, plasmid, mitochondrial, and the yeast genomes (Carles-Kinch and Kreuzer, 1997; Itoh and Tomizawa, 1980; Kreuzer and Brister, 2010; Lee and Clayton, 1998; Masukata and Tomizawa, 1990; Stuckey et al., 2015; Xu and Clayton, 1995). In addition, 76% of R-loop-prone human CpG island promoters function as early, constitutive replication origins (Cadoret et al., 2008; Mukhopadhyay et al., 2014; Picard et al., 2014; Sequeira-Mendes et al., 2009), further illustrating the links between R-loops and

replication origins. Collectively, these studies suggest that R-loops represent a novel type of biological signal that can trigger specific biological outputs in a context-dependent manner (Chedin, 2016).

Early on, R-loops were shown to form at repeated class switch sequences during the process of immunoglobulin class switch recombination (Huang et al., 2007; Huang et al., 2006; Yu et al., 2003). Such programmed R-loops have been linked to the initiation of class switch recombination via the introduction of double-stranded DNA breaks (Yu and Lieber, 2019). Since then, "harmful" R-loop formation, generally understood to result from excessive R-loop formation, has been implicated in various processes linked to genomic instability such as hyper recombination, transcription-replication collisions, replication and transcriptional stress (for review (Costantino and Koshland, 2015; Crossley et al., 2019a; Garcia-Muse and Aguilera, 2019; Santos-Pereira and Aguilera, 2015; Skourti-Stathaki and Proudfoot, 2014)). As a consequence, the accurate mapping of R-loop structures represents an exciting and important challenge to better understand the distribution and function of these structures in health and disease.

DNA:RNA immunoprecipitation (DRIP) relies on high affinity of the S9.6 monoclonal antibody for DNA:RNA hybrids (Boguslawski et al., 1986). DRIP-seq permits robust genome-wide profiling of R-loop formation (Ginno et al., 2012; Sanz and Chedin, 2019b). While useful, this technique suffers from limited resolution due to the fact that restriction enzymes are used to achieve gentle DNA fragmentation. In addition, DRIP-seq does not provide information on the directionality of R-loop formation. Here we report a variant of DRIP-seq that permits the mapping of R-loops at high resolution in a strand-specific manner. This method relies on sonication to fragment the genome prior to immunoprecipitation and the method is thus called sDRIP-seq (sonication DNA:RNA immunoprecipitation coupled to high throughput sequencing) (Figure 2.1). The use of sonication permits an increased resolution and limits restriction enzyme-linked fragmentation biases observed in DRIP-seq approaches (Halasz et al., 2017). sDRIP-seq produces R-loop maps that are in strong agreement with results from both DRIP-seq and the previously described high-resolution DRIPc-seq method in which sequencing libraries are built from the RNA strands of immunoprecipitated R-loop structures (Sanz and Chedin, 2019b).

Faced with a plethora of methods to choose from, users may wonder which particular DRIP-based approach is preferable for their needs. We offer the following advice. DRIP-seq, despite its limitations, is technically easiest and is the most robust (highest yields) of all three methods discussed here; it thus remains broadly useful. Numerous DRIP-seq datasets have been published, which provides a useful comparison point for new datasets. Finally, the bioinformatic analysis pipeline is simpler as the data is not stranded. We recommend that new users begin honing their R-loop mapping skills with DRIP followed by qPCR and DRIP-seq. sDRIP-seq represents a slightly higher degree of technical difficulty: the yields are slightly reduced due to sonication (discussed below) and the sequencing library process is slightly more complex. Yet, the gain of strandedness and higher resolution is invaluable. We note that sDRIP-seq will capture both two-stranded RNA:DNA hybrids and three-stranded R-loops. Due to the library construction steps, DRIP-seq will not capture two-stranded RNA:DNA hybrids. DRIPc-seq is the most technically demanding and requires higher amount of starting materials. In return, it offers the highest resolution and strandedness. Because sequencing libraries are built from the RNA moiety of R-loops or hybrids, DRIPc-seq may suffer from possible RNA contamination, especially since S9.6 possesses residual affinity for dsRNA (Chedin et al., 2021b; Hartono et al., 2018; Phillips et al., 2013). sDRIP-seq permits strand-specific, high resolution mapping without worries about RNA contamination since sequencing libraries are derived from DNA strands. Overall, these three methods remain useful and present differing degrees of complexity and slightly different caveats. All three, however, produce highly congruent datasets (Chedin et al., 2021b) and are highly sensitive to RNase H pre-treatment, which represents an essential control to ensure signal specificity (Sanz and Chedin, 2019b; Smolka et al., 2021b). We note that given the size selection imposed on sequencing libraries, small hybrids (estimated < 75 bp), such as those forming transiently around lagging strand DNA replication priming sites (Okazaki primers) will be excluded. Similarly, since all DRIP methods involve DNA fragmentation, unstable R-loops that require negative DNA supercoiling for their stability will be lost (Stolz et al., 2019a). Thus DRIP approaches may underestimate R-loop loads, especially for short, unstable R-loops that may be best captured using "in vivo" approaches (Chedin et al., 2021b; Sanz and Chedin, 2019b). We note that R-loops can also be profiled in an S9.6-independent manner

at deep coverage, high-resolution and in a strand-specific manner on single DNA molecules after sodium bisulfite treatment (Malig et al., 2020b). Additionally, strategies using a catalytically inactive RNase H1 enzyme have been employed to map native R-loops "in vivo", highlighting short, unstable R-loops that form primarily at paused promoters (Chen et al., 2019; Wang et al., 2021b; Yan and Sarma, 2020).

2.3 Protocol

The following protocol is optimized for the human Ntera-2 cell line grown in culture but it has been successfully adapted without modification to a range of other human cell lines (HEK293, K562, HeLa, U2OS), primary cells (fibroblasts, B-cells) as well as in other organisms with small modifications (mouse, fly).

1. Cell harvest and lysis

1.1. Culture Ntera-2 cells to 75-85% confluency. To start any DRIP procedure, optimal cell count should be 5 to 6 million cells with >90% viable counts.

1.2. Wash cells once with 1x PBS and add 1.5ml of Trypsin-EDTA 1X and incubate for 2 minutes at 37°C until the cells dissociate from the dish.

1.3. Add 5 ml of warm media and, after pipetting well to resuspend cells into a single cell suspension, transfer the content in a new 15 ml tube and gently pellet the cells at 1,000 rpm for 3 minutes.

1.4. Wash the cells once with 5 ml of 1x PBS and gently pellet the cells at 1,000rpm for 3 minutes.

1.5. Fully resuspend the cells in 1.6 mL of TE buffer (10 mM Tris-Cl pH 7.5, 1 mM EDTA pH 8.0). Add 5 µL of proteinase K (20 mg/mL stock solution) and 50 µL of SDS (20% stock solution) and invert gently the tubes 5 times until solution become viscous. Do not try to pipet the solution, only mix by inversion.

1.6. Incubate overnight at 37°C.

2. DNA extraction

2.1. Pour the DNA lysate into a pre-spun 15 mL high density Maxtract phase lock gel tube and add 1 volume (1.6 mL) of Phenol/Chloroform Isoamyl alcohol (25:24:1). Gently invert 5 times and spin down at 1,500 g for 5 minutes.

2.2. Add 1/10 volume of 3M sodium acetate (NaOAc) pH 5.2 and 2.5 volumes of 100% Ethanol to a new 15 mL tube. Pour in the top aqueous phase from the phase lock gel tube and invert softly until the DNA is fully precipitated (up to 10 minutes).

2.3. Spool DNA threads using wide bore 1000 μL tip and transfer to a clean 2 mL tube while taking care of not carrying over residual supernatant.

2.4. Wash the DNA by adding 1.5 mL of ethanol 80% and gently invert the tube 5 times. Incubate for 10 minutes.

2.5. Repeat the previous step twice. Do not centrifuge during the wash steps. Carefully remove as much ethanol as possible by pipetting after the last wash while trying to not disturb DNA.

2.6. Allow the DNA to air dry completely while inverting the tube. This step can take 30 minutes to an hour depending on the amount of DNA.

2.7. Add directly on the DNA pellet 125 μL of TE if you are fragmenting the DNA through restriction enzyme digestion or 100 μL of TE if you are shearing the DNA through sonication. Keep on ice for one hour and gently resuspend DNA by pipetting a few times with a wide bore 200 μL tip. Leave on ice another hour before starting the fragmentation step.

3. DNA fragmentation

For restriction enzyme-based DRIP-seq, please follow step 3.1. For sonication-based DRIP-seq, follow skip to step 3.2.

3.1. Restriction enzyme (RE) fragmentation

3.1.1. Digest the resuspended genomic DNA (very viscous) using a cocktail of REs according to supplier's instructions. Add 0.1 mM spermidine to the final reaction. We advise to use a cocktail of 4 to 5 enzymes with 30U of each enzyme in a total volume of 150 µL. The initial cocktail developed for DRIP-seq (HindIII, SspI, EcoRI, BsrGI, XbaI) (Ginno et al., 2012) was developed to generate an average fragment length of 5 kilobases, avoid any interference with CpG methylation, and spare GC-rich regions of the genome. Other cocktails are possible (Ginno et al., 2013)). These cocktails are suitable for both the human and mouse genomes but can be adjusted as needed. Incubate overnight at 37°C. The DNA mixture post digest should no longer be viscous. Any remaining viscosity at this step is indicative of an incomplete digestion. If observed, add an additional 10U of each enzyme and incubate for another 2-4 hours at 37°C. Note that users may not digest the entire pellet in the event they harvested more cells than recommended here.

3.1.2. Gently pipet the overnight digested DNA (150 µL) into a pre-spun 2 mL phase lock gel light tube. Add 100 µL of water and one volume (250 µL) of Phenol/Chloroform Isoamyl alcohol (25:24:1). Gently invert 5 times and spin down at 16,000 g for 10 minutes.

3.1.3. Add 1.5 µL of glycogen, 1/10 volume 3M NaOAc pH 5.2 and 2.5 volumes 100% Ethanol to a new 1.5 mL tube. Pipet the DNA from the phase lock gel tube and mix by inverting 5 times. Incubate one hour at -20°C.

3.1.4. Spin at 16,000 g for 35 minutes at 4°C. Wash DNA with 200 µL 80% ethanol and spin at 16,000 g for 10 minutes at 4°C.

3.1.5. Air dry the pellet and add 50 µL of TE buffer to the pellet. Leave the tube on ice for 30 minutes and gently resuspend the DNA.

3.1.6. Measure the concentration (OD_{260}) of the fragmented DNA on a Nanodrop or equivalent.

3.1.7. Optional but recommended: load 1 µg of digested DNA on a 0.8% agarose gel alongside a size marker to verify that the digest is complete. If incomplete, additional enzyme can be added. Incomplete digestion can lead to loss of resolution after immunoprecipitation.

3.1.8. After this step, 10 µgs of digested DNA can be treated with 4 µL of NEB ribonuclease H (RNase H) for 1 to 2 hours at 37°C in order to control that the signal retrieved upon immunoprecipitation derives from DNA:RNA hybrids. You can then proceed to S9.6 immunoprecipitation (step 4). Note that digested DNAs can be kept frozen at -80°C for up to one month without significant loss of yield.

3.2. Sonication

3.2.1. Sonicate all or part of the extracted DNA in a 0.5 mL Eppendorf tube in 100 µL total volume. Perform 15 to 20 cycles of 30 sec ON / 30 sec OFF on a Diagenode Bioruptor NGS sonicator (spin after 5, 10 and 15 cycles to ensure homogeneous sonication).

3.2.2. Measure the concentration (OD_{260}) of sonicated DNA on a Nanodrop or equivalent. At this step, the viscosity of the DNA should have disappeared.

3.2.3. Run an agarose gel to confirm the size distribution of sonicated DNA (300-500 bp). Over-sonicating DNA can lead to significant reduction in yield resulting from breakage and dissociation of R-loop structures.

3.2.4. After this step, 10 µgs of sonicated DNA can be treated with 4 µL of NEB RNase H for 1 to 2 hours at 37°C in order to control that the signal retrieved upon immunoprecipitation derives from DNA:RNA hybrids. You can then proceed to S9.6 immunoprecipitation (step 4).

4. S9.6 immunoprecipitation

The immunoprecipitation steps are similar regardless of whether DNA was fragmented through REs or sonication.

4.1. Prepare three tubes and aliquot 4.4 µg of fragmented DNA in a final volume of 500 µL of TE per tube. Save 50 µL (1/10 of the volume) from each tube to use later as an input.

4.2. Add 50 µL of 10X binding buffer (100 mM NaPO4 pH 7, 1.4 M NaCl, 0.5% Triton X-100) and 10 µL of S9.6 antibody (1 mg/ml) to the 450 µL of diluted DNA.

4.3. Incubate overnight at 4°C on a mini-tube rotator at 7 to 10 rpm.

4.4. For each tube, wash 50 µL of Protein A/G agarose bead slurry with 700 µL of 1X binding buffer by inverting the tubes on a mini-rotator at 7 to 10 rpm at room temperature for 10 minutes. Spin down the beads at 1,100 g for one minute and discard supernatant. Repeat this step once.

4.5. Add the DNA from step 4.3 to the 50 µL of beads and incubate for 2 hours at 4°C while inverting at 7 to 10 rpm on a mini-rotator.

4.6. Spin down the beads one minute at 1,100 g and discard supernatant.

4.7. Wash the beads with 750 µL of binding buffer 1X by inverting at 7 to 10 rpm on a mini-rotator for 15 minutes. Spin down one minute at 1,100g and discard supernatant. Repeat this step once.

4.8. Add 250 µL of elution buffer (50 mM Tris-Cl pH 8, 10 mM EDTA pH8, 0.5% SDS) and 7 µL of proteinase K (20 mg/mL stock) to the beads and incubate with rotation at 55°C for 45 minutes.

4.9. Spin down the beads one minute at 1,100 g. Transfer the supernatant to a pre-spun 2 mL phase lock gel light tube and add one volume (250 µL) of Phenol/Chloroform Isoamyl alcohol (25:24:1). Invert tubes 5 times and spin down for 10 minutes at 16,000 g at room temperature.

4.10. Add 1.5 µL of glycogen, 1/10 volume 3M NaOAc pH 5.2 and 2.5 volumes of 100% Ethanol to a new 1.5 mL tube. Pipet the DNA from the phase lock gel tube and mix by inverting 5 times. Incubate one hour at -20°C.

4.11. Spin at 16,000 g for 35 minutes at 4°C. Wash DNA with 200 µL 80% ethanol and spin at 16,000 g for 10 minutes at 4°C.

4.12. Air dry pellets and add 15 µL of 10 mM Tris-Cl pH 8 in each tube. Leave tubes on ice for 20 minutes and gently resuspend. Combine the 3 tubes in one (45 µL).

4.13. Check DRIP efficiency by qPCR using 5 µL of the 45 µL resuspended DNA (see representative results). Dilute the 5 µL in 10 µL of water and use 2 µL per reaction.

5. Pre-library step for sonicated DNA only

Sonication leads the displaced ssDNA strand of R-loops to break. Thus, three-stranded R-loop structures are converted into two-stranded DNA:RNA hybrids upon sonication. As a result, these RNA:DNA hybrids must be converted back to double-stranded DNA prior to library construction. Here we employ a second strand synthesis step. An alternative approach that has been successfully used is to instead perform a single-stranded DNA ligation followed by a second strand synthesis (Crossley et al., 2020).

5.1. To the 40 µL of DRIP'ed DNA from step 4.12, add 20 µL of 5X second strand buffer (200 mM Tris pH 7, 22 mM MgCl$_2$, 425 mM KCl), 10 mM dNTP mix (dATP, dCTP, dGTP, and dTTT or **dUTP if you are planning to achieve strand-specific DRIP sequencing**), 1 µL 16 mM NAD and 32 µL water. Mix well and incubate 5 minutes on ice.

5.2. Add 1 µL of DNA polymerase I (10 units), 0.3 µL of RNase H (1.6 units) and 0.5 µL of *E. coli* DNA ligase. Mix and incubate at 16°C for 30 minutes.

5.3. Immediately clean up the reaction using Ampure beads with a ratio of 1.6X. Elute DNA in 40 µL of 10 mM Tris-Cl pH8.

6. Pre-library sonication step for RE DNA only.

DRIP leads to the recovery of RE fragments that are often kilobases in length and thus not suited for immediate library construction. To reduce the size of the material for library construction, sonicate the immunoprecipitated DNA in a 0.5 mL Eppendorf tube. Perform 12 cycles of 15 sec ON / 60 sec OFF on a Diagenode Bioruptor NGS sonicator (spin after 6 cycles to ensure homogeneous sonication). Proceed to step 7.

Optional step: the immunoprecipitated material still carries three-stranded R-loops which respond to sonication differently than the flanking double-stranded DNA. To even out DRIP profiles, we recommend treating the immunoprecipitated material with 1 µL of NEB RNase H in 1x RNase H buffer for 1 hour at 37°C prior to sonication.

7. Library construction

7.1. Perform end repair by adding to the 40 µL from step 4.12 (RE fragmentation) or step 5.3 (sonication shearing) 5 µL of NEB 10X end repair module buffer, 2.5 µL of 10 mM ATP and 2.5 µL NEBNext End repair module enzyme (50 µL total). Mix well and incubate for 30 minutes at room temperature. Also include 1 mg of RE-digested and sonicated (DRIP) or sonicated (sDRIP) input DNA to create control sequencing libraries corresponding to the input DNA.

7.2. Clean up the reaction using AMPure beads (1.6X ratio) and elute in 34 µL of 10 mM Tris-Cl pH8.

7.3. Perform A-tailing by adding 5 µL of NEB buffer 2, 10 µL of 1 mM dATP and 1 µL of NEB Klenow exo- (50 µL total). Mix well and incubate for 30 minutes at 37°C.

7.4. Clean up the reaction using AMPure beads (1.6X ratio) and elute in 12 µL of 10 mM Tris-Cl pH8.

7.5. Ligate adapters by adding 15 µL of NEB 2X quick ligation buffer, 1 µL of 15 µM Illumina adapters and 2 µL of NEB quick ligase (30 µL total). Mix well and incubate for 20 minutes at room temperature.

7.6. Clean up the reaction using AMPure beads (1X ratio) and elute in 20 µL of 10 mM Tris-Cl pH8.

7.7. If you used sonication shearing and dUTP in step 5.1 and want a strand-specific DRIP, add 1.5 µL (1.5 U) of AmpErase Uracil N-glycosylase. Incubate 30 minutes at 37°C.

7.8. PCR amplify 10 µL of the library from step 6.6 or 6.7. Add 1 µL of PCR primer 1.0 P5, 1 µL of PCR primer 2.0 P7, 15 µL of Phusion master mix and 3 µL of water. Mix well.

7.9. In a thermo cycler, run the following program:

Cycle number	Duration	Temperature
1	30 sec	98°C
2-15	10 sec	98°C
	30 sec	60°C

	30sec	72°C
16	5 min	72°C
17	Hold	12°C

7.10.   Proceed to a two-step clean-up of your library using AMPure. Use first a ratio of 0.65X to

remove fragments over 500 bp. Keep supernatant. Proceed to a 1X ratio on the supernatant to remove

fragments under 200 bp. Elute in 12 µL.

8. Quality control

8.1. Check R-loop enrichments with qPCR on two negative and three positive loci using the Pfaffl method

using 1 µL of the clean-up library from step 6.10. Dilute 1 µL of the library in 10 µL of water and use

2 µL per locus.

8.2. Check the size distribution of your cleaned-up library from step 6.10 using an Agilent High

sensitivity DNA 1000 kit.

2.4 Representative results

DRIP as well as sDRIP can be analyzed through qPCR (Figure 2.2A) and/or sequencing (Figure

2.2B). After the immunoprecipitation step, the quality of the experiment must be first confirmed by qPCR

on positive and negative control loci, as well as with RNase H-treated controls. Primers corresponding to

frequently used loci in multiple human cell lines are provided in Table 1. The results from qPCR should be

displayed as a percentage of input, which corresponds to the percentage of cells carrying an R-loop at the

time of the lysis for a given locus. In a successful DRIP experiment, the yield for negative loci should be

less than 0.1% whereas positive loci can vary from 1% to over 10% for highly transcribed loci like *RPL13A*

(Figure 2.2A). For sDRIP, yields are typically lower (20-50%) as judged by DRIP-qPCR but appear to

affect recovery uniformly such that no particular subset of R-loops is affected more than another. As a

result, maps derived from DRIP, sDRIP and DRIPc are in good agreement (Figure 2.2B). qPCR data can also be displayed as fold enrichment of the percentage of input for positive loci over negative loci, thus assessing the specificity of the experiment. Fold enrichments typically range from a minimal of 10-fold to over 200-fold depending on the loci chosen for analysis. When precise quantification across multiple samples representing gene knockdowns, knockouts, or various pharmacological treatments, is required, the use of spike in controls to normalize inter-sample experimental variation is highly encouraged. Such spike-ins can correspond to synthetic hybrids (Crossley et al., 2020) or genomes of unrelated species (Svikovic et al., 2019).

DRIP and sDRIP materials can be sequenced using single or paired-end sequencing strategies. Data can be extracted and analyzed in a similar manner as most ChIP data using standard computational pipelines (see (Sanz and Chedin, 2019b) for DRIP-relevant information). After adapter trimming and removal of PCR duplicates, reads can be mapped to a reference genome and uploaded to a genome browser. A typical expected output of DRIP and sDRIP is shown in Figure 2.2B. The DRIP output is represented by the only green track as it does not allow strand specificity whereas sDRIP shows R-loop mapping to the positive and negative strands indicated respectively in red and blue. Control tracks corresponding to a sample pre-treated with RNase H show a clear reduction of signals, confirming the specificity of the technique for RNA:DNA hybrid-derived materials. The gains in resolution permitted by sDRIP are clearly illustrated when comparing the sizes of input DNA material (Figure 2.2C). The reproducibility of sDRIP-seq, along with the global impact of RNase H1 pre-treatment and the correlation between sDRIP-seq and DRIPc-seq are depicted by XY plots in Figure 2.2D.

2.5 Discussion

We describe here two protocols to map R-loop structures in potentially any organism using the S9.6 antibody. DRIP-seq represents the first genome-wide R-loop mapping technique developed. It is an easy, robust, and reproducible technique that allows one to map the distribution of R-loops along any

genome. The second technique, termed sDRIP-seq, is also robust and reproducible but achieves higher resolution and strand-specificity owing to the inclusion of a sonication step and a stranded sequencing library construction protocol. Both techniques are highly sensitive to RNase H treatment prior to immunoprecipitation, confirming that the signal is principally derived from genuine RNA:DNA hybrids. Finally, when comparing immunoprecipitation yields between R-loop positive and R-loop negative loci, both techniques offer up to a 100-fold difference in several human cell lines, providing high specificity mapping with low background.

When considering which method to implement, it is useful to consider their respective strengths and limitations. As previously noted, DRIP-seq produces maps with a lower resolution and does not give information on the strandedness of R-loop formation. The lower resolution is mainly a product of the use of REs to fragment the genome. This gentle method is best at preserving R-loops, thereby allowing unsurpassed recovery of such structures, and making DRIP-seq very robust. To circumvent the issue of limited resolution while preserving high recovery, RE cocktails can be adapted and/or maps resulting from different RE cocktails can be combined to improve resolution (Ginno et al., 2013). A technique using 4 bp cutters has been developed to improve the resolution of DRIP-seq and may achieve strand-specific mapping (Xu et al., 2017; Yang et al., 2019), although the resulting datasets have not yet been systematically compared to other human datasets. It is important to note that in RE-based approaches, larger fragments tend to be recovered more efficiently because they can carry multiple R-loop forming regions. This bias must be taken into account when analyzing DRIP-seq datasets. Similarly, peak calling for DRIP-seq data must be ultimately translated into R-loop-positive RE fragments, since it is these fragments that are immunoprecipitated and the position of R-loops within these fragments can't be inferred. In general, we recommend that users first adopt RE-based DRIP-seq to learn the method and build their confidence in achieving the yields documented in Figure 2.2A. sDRIP-seq typically results in lower yields, which could result in maps with lower signal to noise ratios in untrained hands. The use of sonication as a means of fragmenting the genome offers in return a great improvement in resolution since the non-R-looped portions

that typically constitute the majority of RE fragments will be broken off, allowing S9.6 to principally retrieve the R-looped portions (Figure 2.1). The additional benefit of strand-specificity provides numerous further benefits to the understanding of R-loop formation mechanisms, making sDRIP-seq a method of choice for the study of R-loops.

Importantly, maps obtained via DRIP-seq and sDRIP-seq represent the average distribution of R-loops through a cell population; thus, the length and position of individual R-loops cannot be addressed with those techniques. For this, an independent and complementary method termed single-molecule R-loop footprinting (SMRF-seq) (Malig et al., 2020b) can be leveraged to reveal individual R-loops at high-resolution in a strand-specific manner. Assessment of R-loop formation using SMRF-seq over 20 different loci, including independently of S9.6, revealed a strong agreement between collection of individual R-loop footprints and the population average distribution gather by DRIP-based approaches (Malig et al., 2020b), lending strong support to DRIP-based approaches. It is also important to consider that R-loop mapping data only provides a snapshot of R-loop genomic distribution and does not provide information on the dynamics of R-loop formation, stability, and resolution. DRIP approaches, combined with specific drug treatments and an evaluation of R-loop distributions through time series, can nonetheless be deployed to address these parameters (Crossley et al., 2020; Sanz et al., 2016). The limitations of R-loop profiling methodologies are particularly important to keep in mind when the goal is to characterize altered R-loop distributions in response to a genetic, environmental, or pharmacological perturbations. In addition to those already described above, it is key to consider any possible change to nascent transcription since these will inherently cause R-loop changes owing to the co-transcriptional nature of these structures. These issues and guidelines for developing rigorous R-loop mapping approaches have been extensively discussed (Chedin et al., 2021b; Vanoosthuyse, 2018) and readers are encouraged to refer to these studies.

2.6 Figures
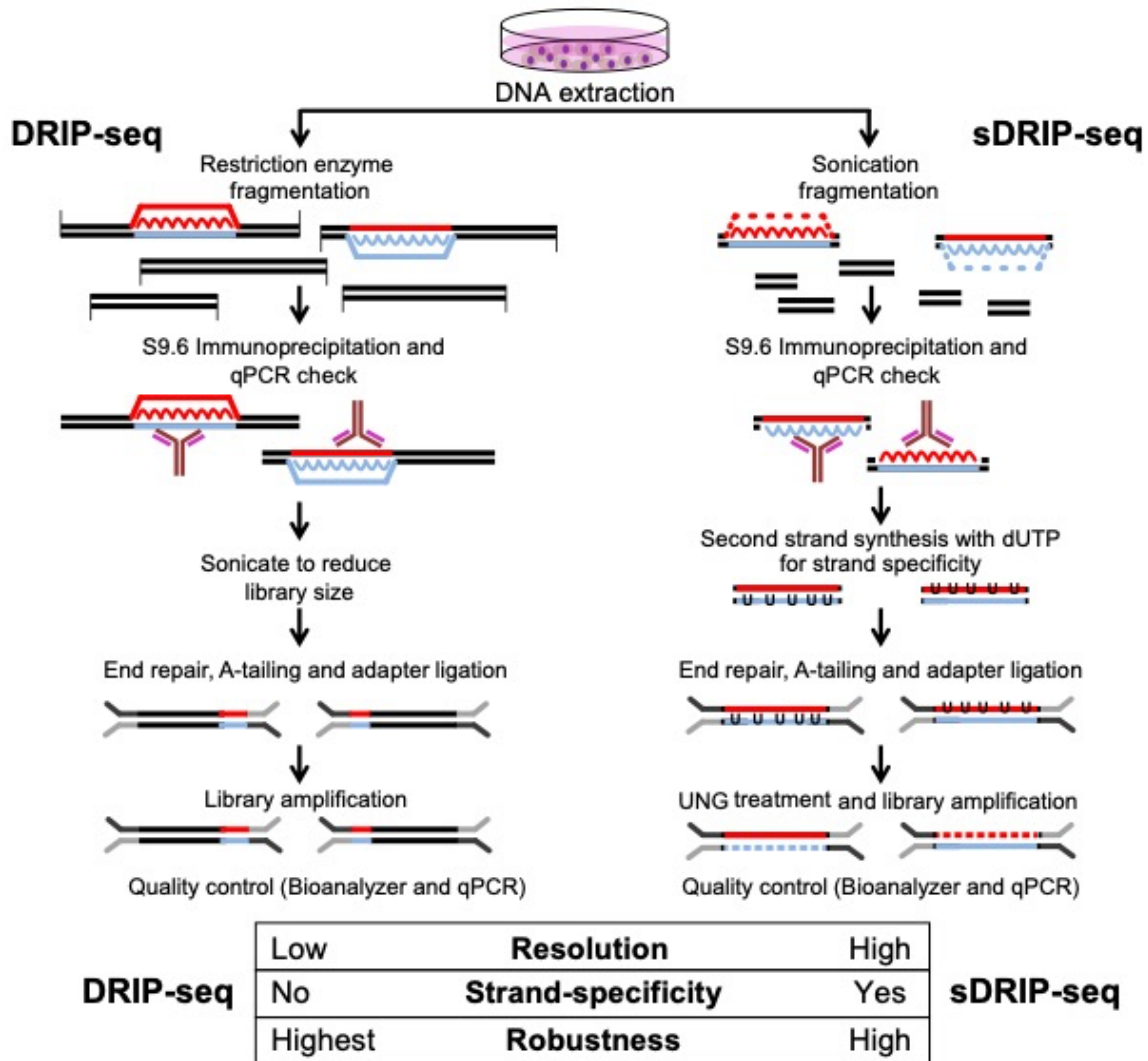
Figure 2.1: Overview of the DRIP-seq and sDRIP-seq procedures. Both approaches start by the same DNA extraction steps developed to preserve R-loops (RNA strands within R-loops are represented by squiggly lines). For DRIP-seq, the genome is fragmented using restriction enzymes, often resulting in kilobase-size fragments within which shorter R-loops are embedded. For sDRIP-seq, the genome is fragmented via sonication which results in smaller fragments and the shearing and loss of the displaced single-strand of R-loops (indicated by dashed lines). Following immunoprecipitation with the S9.6 antibody, DRIP leads to the recovery of three-stranded R-loops embedded within restriction fragments, while sDRIP recovers two-stranded RNA:DNA hybrids with little flanking DNA, ensuring higher

resolution. For sDRIP, a library construction step must be included to convert RNA:DNA hybrids back to duplex DNA. As shown here, this is an opportunity to build strand-specific libraries. As detailed in the protocol itself, exogenous treatment with RNase H represents a key control for the specificity of both procedures; they are not shown here.
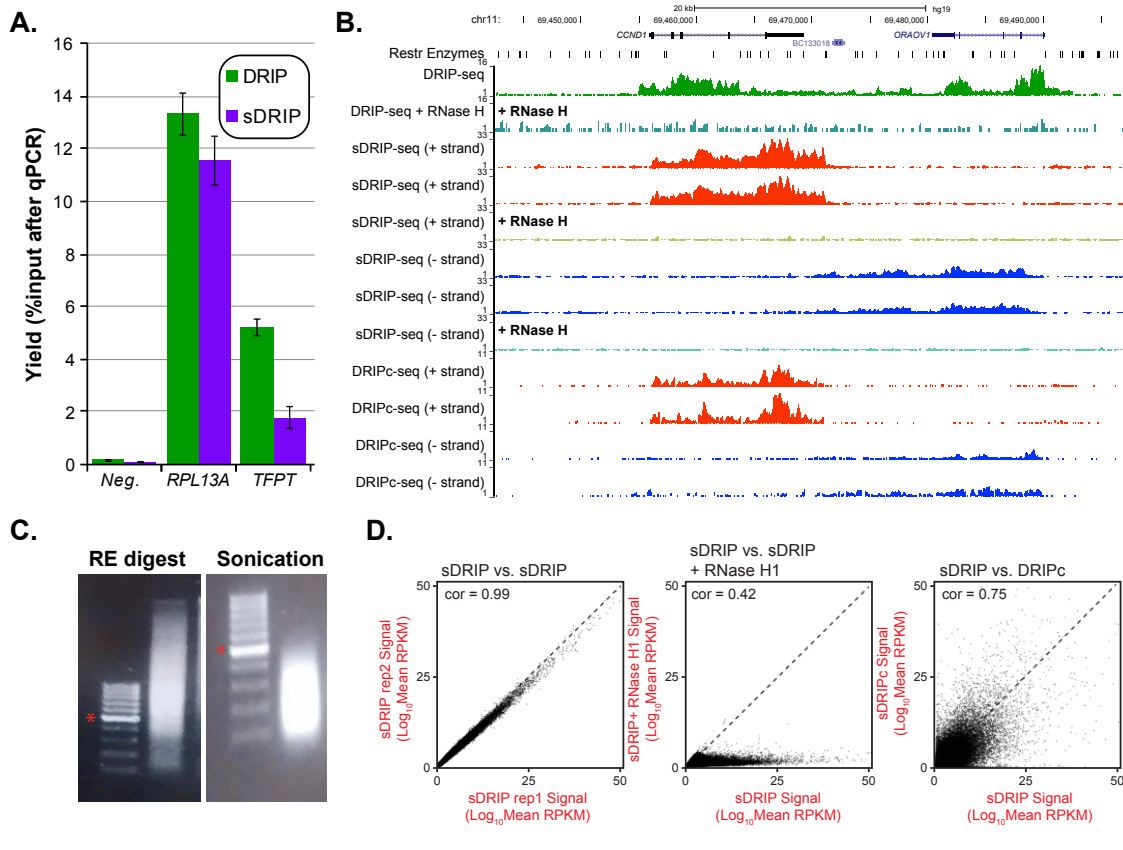
Figure 2.2: Result of R-loop mapping strategies. A. qPCR results from successful immunoprecipitations using the DRIP and sDRIP method (corresponding to qPCR check step 4.13). Results are from two independent experiments from human Ntera2 cells at a negative locus and two positive loci, including the highly R-loop-prone *RPL13A* locus and the moderately R-loop-prone locus *TFPT*. The y-axis indicates the yield of the immunoprecipitation as a percentage of the input DNA. Note that the recovery is slightly more robust for DRIP than sDRIP. **B.** The results of R-loop mapping conducted in human Ntera-2 cells are shown

over a region centered around the *CCND1* and neighboring *ORAOV1* genes. The first two tracks correspond to DRIP-seq results, without and with RNase H treatment, respectively. The position of the restriction enzymes used to fragment the genome are shown at the top. The next six tracks represent the results of strand-specific sDRIP-seq, broken down between (+) and (-) strands (two replicates each) and pre-treated with RNase H, or not, as indicated. The last four track represents the results of R-loop mapping via the high-resolution strand-specific DRIPc-seq method (Sanz *et al.*, 2016; Sanz and Chedin, 2019), where libraries are built from the RNA strands of R-loops. As can be clearly seen, the *CCND1* and *ORAOV1* genes lead to R-loop formation on the (+) and (-) strands, respectively, consistent with their directionality. RNase H treatment abolishes signal, as expected. **C**. Input DNA materials after restriction enzyme fragmentation (left) and sonication (right) are shown after the materials were separated by agarose gel electrophoresis. The DNA ladder corresponds to a 100 bp ladder and the 500 bp band is highlighted by an asterisk. **D**. XY signal correlation plots are shown to illustrate the reproducibility of sDRIP-seq (left), the overall sensitivity of sDRIP-seq to RNase H1 pre-treatment (middle), and the global correlation between sDRIP-seq and DRIPc-seq (right). All data from Ntera-2 human cells.

2.7 Author Contributions

L.A.S performed all experiments with the assistance of D.C.G. LA.S and D.C.G will be involved in the filming of the protocol and writing the script for the video protocol. F.C. conceived and supervised the project, analyzed data, and wrote the manuscript along with all authors.

CHAPTER III: SF3B1-targeted Splicing Inhibition Triggers Global Alterations in Transcriptional Dynamics and R-Loop Metabolism

This chapter has been modified from the manuscript that was published under the same title:

SF3B1-targeted Splicing Inhibition Triggers Global Alterations in Transcriptional Dynamics and R-Loop Metabolism. Daisy Castillo-Guzman, Stella R. Hartono, Lionel A. Sanz, Frédéric Chédin. bioRxiv 2020.06.08.130583; doi: https://doi.org/10.1101/2020.06.08.130583

This manuscript has been modified to match the formatting of the dissertation. My contributions to the original manuscript was executing and analysis of all experiments. Hartono S. R. developed the computational pipeline and performed the pipeline analysis. This manuscript will be resubmitted for publication at a peer-reviewed journal.

3.1 Abstract

Efficient co-transcriptional splicing is thought to suppress the formation of genome-destabilizing R-loops upon interaction between nascent RNA and the DNA template. Inhibition of the SF3B splicing complex using Pladienolide B (PladB) in human K562 cells caused widespread intron retention and nearly 2,000 instances of R-loops gains. However, only minimal overlap existed between these events, suggesting that unspliced introns by themselves do not cause excessive R-loops. R-loop gains were instead driven by readthrough transcription at a subset of stress-response genes, defining a new class of aberrant "downstream of genes" (DoG) R-loops. DoG R-loops formed early compared to the kinetic response of DNA damage and showed limited increase in γH2AX. The increase of γH2AX was also not limited to DoG R-loops. Thus DoG R-loops were temporally and spatially uncoupled from loci experiencing DNA damage. Unexpectedly, the predominant response to splicing inhibition was a global R-loop loss resulting from accumulation of promoter-proximal paused RNA polymerases and defective elongation associated with premature termination. Thus, SF3B1-targeted splicing inhibition triggered profound alterations in transcriptional dynamics, leading to unexpected disruptions in the global R-loop landscape.

3.2 Introduction

During transcription, the nascent RNA can anneal to the DNA template strand behind the advancing RNA polymerase (RNA Pol), forming a stable RNA:DNA hybrid and causing the non-template DNA strand to loop out. The resulting non-B DNA structure, called an R-loop, is facilitated by favorable DNA sequence and negative superhelicity (Chedin and Benham, 2020; Drolet et al., 2003; Stolz et al., 2019a). R-loops have been described from bacteria to plants to mammals (Santos-Pereira and Aguilera, 2015) and represent a prevalent class of alternative DNA structures (Chedin, 2016). Studies over the last decade have implicated R-loop formation in a variety of physiological processes such as transcription termination (Proudfoot, 2016; Sanz et al., 2016; Skourti-Stathaki et al., 2011), chromatin patterning and gene expression control (Chedin, 2016; Niehrs and Luke, 2020; Sanz et al., 2016; Skourti-Stathaki et al., 2019; Tan-Wong et al., 2019), as

well as class switch recombination (Yu et al., 2003; Yu and Lieber, 2019). By contrast, so-called aberrant R-loops formed under a variety of pathological conditions have been extensively linked to phenomena of genome instability (Crossley et al., 2019a; Garcia-Muse and Aguilera, 2019; Hamperl and Cimprich, 2014). Not surprisingly, a number of human diseases have recently been linked to altered R-loop metabolism (Richard and Manley, 2017). Despite the rising importance of R-loops in adaptive and maladaptive processes, many questions remain regarding the manner by which they are formed and regulated.

It is widely accepted that R-loops form co-transcriptionally upon re-invasion of the nascent RNA *in cis*. Interestingly, alterations in co-transcriptional processes such as splicing, RNA export, cleavage and polyadenylation, have been linked to phenomena of R-loop-mediated genome instability from yeast to human cells (Aguilera, 2005b; Aguilera and Garcia-Muse, 2012; Chan et al., 2014; Huertas and Aguilera, 2003; Li and Manley, 2005, 2006; Paulsen et al., 2009; Stirling et al., 2012). Earlier work suggested that inactivation in the splicing regulator SRSF1 triggered the accumulation of R-loops, presumably upon formation of excessive R-loops *via* newly unspliced transcript portions (Tresini et al., 2015). SRSF1-depleted cells were also reported to show the accumulation of double-stranded DNA breaks (DSBs), a hypermutagenic phenotype, G2 arrest, and loss of cell viability (Li and Manley, 2005). Importantly, these phenotypes could be reversed upon over-expression of ribonuclease H1 (RNase H1), an enzyme with the clear biochemical ability to degrade RNA in the context of RNA:DNA hybrids (Cerritelli and Crouch, 2009). This suggested that excessive R-loop formation is directly involved in driving genome instability, and that splicing factors or efficient splicing and processing can limit R-loop formation. Systematic genetic screens in human cells identified splicing-related genes as the top category of factors involved in the prevention of R-loop-mediated genome instability, as scored by the phosphorylated H2AX (γH2AX) DNA damage marker (Paulsen et al., 2009). Complementary work in yeast further underscored that introns attenuate R-loop formation and transcription-associated genetic instability *via* the recruitment of the spliceosome onto the pre-mRNA (Bonnet et al., 2017). Altogether, these studies suggest that proper splicing, and packaging of the nascent RNA into ribonucleoprotein particles for efficient nuclear export

contribute significantly to the maintenance of genome stability by keeping the RNA away from the DNA template and avoiding deleterious R-loop formation.

The nature and distribution of excessive R-loops expected to result from the disruption of co-transcriptional RNA processing have not been established at genome scale. Thus, the spatial relationship between putative aberrant R-loop gains and events of genome instability has never been directly tested. To address these gaps, we focused on splicing as a key co-transcriptional process linked to R-loop regulation. SF3B1 is a conserved and essential core subunit of the SF3B complex, a key component of the U2 spliceosome involved in branch site recognition and selection during pre-mRNA splicing (Sun, 2020). SF3B1 can be specifically inhibited *via* pharmacological compounds, including Pladienolide B (PladB) (Cretu et al., 2018; Kotake et al., 2007; Yokoi et al., 2011). PladB and other macrolide-type splicing inhibitors rapidly abrogate splicing, causing a vast increase in events of intron retention (IR) (Boswell et al., 2017; Carvalho et al., 2017; Kashyap et al., 2015). PladB treatment provides a unique opportunity to determine whether unspliced introns can drive excessive R-loop formation and associated DNA damage, which is expected if the newly retained intron sequences form genome-destabilizing R-loops. This pharmacological approach further allowed us to temporally resolve the resulting changes in R-loop distribution, splicing events, and transcriptional dynamics, providing a unique view into the cellular response to acute splicing inhibition.

3.3 Results

PladB treatment causes broad intron retention and R-loop reduction in human K562 cells.

We first characterized the PladB response by RT-qPCR at the known PladB target *DNAJB1* (Kashyap et al., 2015). IR gradually accumulated over time and could be detected as soon as 15 minutes after treatment, with greatest increase detected at 2 and 4 hours after drug application (Figure 3.S1A, B). We then used RNA-seq to characterize alterations in splicing patterns resulting from PladB treatment two and four hours after drug application. Human K562 cells either mock-treated with DMSO or untreated were used as controls and two biological replicates were analyzed at each time point (Figure 3.1A). As expected,

PladB treatment caused broad, time-dependent, splicing alterations dominated by events of intron retention (IR) and to a lesser extent skipped exons (SE) (Figure 3.1B). Validation of RNA-seq analysis confirmed that IR was observed at *DNAJB1* in strand-specific RNA-seq analysis of poly(A)-tailed mRNA pools two- and four-hours post treatment (Figure 3.1C). In total, over 7,000 independent IR events were identified in human K562 cells four hours post PladB treatment. Such retained introns provide an ideal cohort to test the hypothesis that splicing inhibition causes gains of R-loops over regions now associated with unspliced introns. To determine how R-loops respond to PladB-mediated splicing inhibition, we profiled R-loop distribution using DRIP-seq (Sanz and Chedin, 2019b) at the same time points. Only a modest number of R-loop gains (RLGs) were observed post PladB treatment with 1,467 and 1,878 peaks identified within the limits of our statistical thresholds two and four hours post-PladB, respectively (Figure 3.1D). Unexpectedly, the vast majority of significant R-loop changes corresponded to events of R-loop loss, highlighting 18,150 and 28,844 peaks two and four hours post-PladB, respectively. In addition, many loci showed no significant R-loop gains or losses.

To determine whether IR associates with RLGs, we first focused on *DNAJB1* that showed clear intron retention: no significant RLGs could be observed (Figure 3.1E). R-loops were in fact significantly reduced at the *RPL13A* gene despite clear IR (Figures 3.S1C, 3.S1D). More broadly, less than 2% of loci with IR events intersected with RLG peaks (121 / 7289; Figure 3.1F). This suggests that contrary to expectations, large-scale intron retention caused by acute PladB-mediated splicing inhibition, did not cause R-loop gains by unspliced transcripts.

Splicing inhibition is associated with readthrough transcription and accompanying R-loops at a subset of genes.

To understand the nature of the RLGs caused by PladB treatment, we asked where these peaks mapped in the genome. 77% of R-loop peaks in untreated K562 cells mapped to genic regions, as expected (Sanz et al., 2016). By contrast, nearly two thirds of the RLGs in PladB-treated cells mapped to intergenic regions (Figure 3.2A), indicating that PladB treatment unexpectedly led to RLGs in non-genic regions. To

assess how such gains arose, we focused on a region with numerous clustered RLGs peaks located downstream of the *VAPA* gene (Figure 3.2B). The peaks extended from the 3′-end of *VAPA* and grew directionally with transcription in a time-dependent manner. After two hours, the edge of the wave of RLGs extended 135 kilobases (kb) downstream of the *VAPA* gene. After four hours, a trail of R-loops was detected 210 kb downstream of *VAPA*. DRIP-qPCR on three additional replicates confirmed that R-loops increased 20-fold 60 kb downstream of the *VAPA* poly(A) site (PAS), rising from background levels to a level matching that of R-loops observed towards the end of the gene itself (Figure 3.2C). When analyzed 200 kb downstream of *VAPA*, R-loops rose 6-fold four hours post treatment. All instances of R-loops were sensitive to RNase H pre-treatment, as expected. This pattern is consistent with a directional and time-dependent propagation of a wave of co-transcriptional R-loops from the *VAPA* gene. Analysis of R-loops using a high-resolution, strand-specific, iteration of DRIP-seq (sDRIP-seq; see methods) confirmed that the RLGs were strand-specific with the *VAPA* template strand and co-directional with *VAPA* transcription (Figure 3.S2A). Similar findings were observed for two additional representative genes, *RPL9* and *CYCS* that showed a strand-specific increase in R-loop signal extending downstream of their PAS (Figure 3.S2B-D).

Given that R-loops form co-transcriptionally, these observations suggest that PladB-mediated splicing inhibition caused readthrough transcription downstream of *VAPA*. RT-qPCR analysis of total RNA confirmed the existence of transcripts downstream of *VAPA* upon PladB treatment (Figure 3.2D). Transcripts were detected 15 and 60 kb downstream of *VAPA* 30 min post PladB treatment and steadily built over time up to a 35- to 60-fold increase over controls. Far downstream of *VAPA* (120 kb), transcripts were only detected 120 and 240 min after treatment, consistent with the steady progression of the RNA polymerase (RNA Pol) machinery from the 3′-end of *VAPA* and not from spurious transcription initiation in intergenic regions. To further explore the possibility that PladB treatment triggers downstream of genes (DoG) transcription, we analyzed nascent transcription in control and PladB-treated K562 cells using EU labeling and RNA sequencing of labeled transcripts. Nascent transcripts were clearly observed downstream of *VAPA* and progressed unidirectionally as a function of time from the normal *VAPA* termination region

(Figure 3.2E). Similar findings were observed at the *CYCS* and *RPL9* genes (Figure 3.S2E-F). Thus, PladB treatment triggered readthrough transcription and accompanying co-transcriptional R-loops.

DoG transcription affects a specific subset of stress-responsive genes and is associated with broad *de novo* R-loop gains.

To determine how broad the PladB-induced transcriptional readthrough was, we systematically annotated whether events of RLGs correspond to DoG R-loops initiating from a neighboring upstream gene. A total of 429 genes, referred to thereafter as DoG genes, were identified with these characteristics, corresponding to the large majority of RLGs (1,505/1,878 peaks; Figure 3.S3A). Metaplots confirmed increased R-loops and nascent transcription extending downstream of the host genes PASs for a distance of 50-75 kb (Figure 3.3A, B; Figure 3.S3B). Annotation of individual DoG R-loops confirmed that the median length of DoG R-loop regions was 49.2 and 50 kb two and four hours post PladB treatment, respectively (Figure 3.3C). This was significantly higher than the length of terminal R-loops recorded for the same genes under control conditions. Interestingly, PladB-sensitive DoG genes showed longer terminal R-loop regions than expression-matched genes even under control conditions (31.5 kb *versus* 18.2 kb) indicating that DoG genes may have an intrinsic propensity to terminate further downstream of their PASs (Figure 3.3C). Collectively, DoG R-loops observed four hours after PladB treatment covered 2.81 megabases of genomic space, defining the first example of a class of excessive R-loops identified at genome-scale.

DoG genes showed significant enrichments for specific gene ontologies (GOs), including translation initiation and ribosome biogenesis, viral process, response to stress, mitochondrial gene expression, and chromatin organization (Figure 3.3D, Figure 3.S3C). The appearance of stress-related and viral-related GOs is notable given precedents for DoG transcription triggered upon environmental stresses and viral infections (Bauer et al., 2018; Cardiello et al., 2018; Hennig et al., 2018; Vilborg et al., 2015; Vilborg et al., 2017). To investigate if DoG genes arise in part due to elevated transcription, we measured changes to mRNA pools following PladB treatment using RNA-seq. 721 genes showed significant increases

in gene expression (Figure 3.3E). These up-regulated genes were dramatically enriched for a few, highly congruent, biological functions, including protein translation and ribosome biogenesis, RNA metabolic processes including RNA splicing, and mitochondrial ATP synthesis (Figure 3.3F). Similar ontologies were observed for DoG genes (Figure 3.3D), suggesting that R-loop gains reflected at least in part an increase in gene expression. Another 1,180 genes were significantly down-regulated and displayed enrichment for functions related to gene expression regulation, cell cycle control, protein ubiquitination, regulation of viral release, and the cellular response to DNA damage (Figure 3.3F). Overall, PladB triggered a strong cellular stress response associated with DoG transcription and accompanying DoG R-loops for a subset of stress-responsive genes.

Global R-loop losses are caused by PladB-induced negative feedback on transcription elongation.

The overwhelming majority of R-loop changes in response to PladB (28,844 / 30,713 events, or 93.8%, 4 hours post PladB) unexpectedly corresponded to R-loop losses (Figure 3.1D). To understand the mechanisms driving these events, we first asked where R-loop losses (RLLs) occurred. Contrary to RLGs, RLLs matched primarily to transcribed gene bodies (44%) and terminal genic regions (25%) (Figure 3.4A). An additional 20% of RLLs matched to normally transcribed regions located immediately downstream of terminal PAS sites. By contrast, only a small minority of RLLs mapped to promoter and promoter-proximal regions. To further understand how these losses originated, we focused on the representative R-loop-prone *DGCR2* gene (Figure 3.4B). RLLs were clearly visible in the gene body and gene terminus, with little initial impact around the promoter region and early first intron. These RLLs were validated by DRIP-qPCR in three additional independent replicates, which confirmed that early R-loops were unchanged but R-loop formation along the gene body was progressively reduced from 10-fold in the middle of the gene to 50-fold by the gene terminus (Figure 3.4C). The simplest hypothesis to account for these observations is that transcription elongation was impaired, leading to a directional reduction of co-transcriptional R-loops. To test this, we took advantage of our matched EU-seq datasets and observed that nascent transcription along *DGCR2* was significantly affected, with EU incorporation only located in the early portion of intron 1,

where R-loops still occurred (Figure 3.4D). Similar observations were made for other genes and further validated by DRIP-qPCR and sDRIP-seq (Figure 3.S4A-D).

We systematically annotated genes showing significant RLL in their gene body or terminal regions and identified 5,651 genes, referred to thereafter as RLL genes. As expected from single gene examples, metaplot analyses confirmed loss of R-loops gradually occurred along gene bodies and was maximal around gene termini (Figure 3.4E; Figure 3.S4E). Clear loss of nascent transcription was also observed along these genes (Figure 3.4F). In general, nascent transcription and R-loops were in strong agreement: 77% of genes independently annotated as showing loss of nascent transcription also showed RLL. The remainder of genes showing EU-seq loss without accompanying RLL simply had little to no R-loop signal under control conditions. Conversely, about half of the RLL genes also showed loss of nascent transcription. Overall, this data suggests that splicing inhibition triggered profound negative feedback on transcription elongation, resulting in a global reduction of co-transcriptional R-loops.

Our findings thus far indicate that PladB treatment resulted in contrasting alterations in transcriptional dynamics for different gene sets. A large subset of RLL genes showed transcription elongation and R-loop losses in response to PladB. By contrast, a smaller group of DoG genes showed readthrough transcription and associated R-loop gains. In addition, we identified a third large class of transcribed genes (n=5,455) that showed no significant loss of gene body R-loops, nor gains of DoG R-loops. These genes will be referred to as "unchanged" and represent a useful internal control. We next sought to identify the mechanisms leading to transcription elongation loss and asked if the gene classes identified above possess differential characteristics that could account for their unique responses to PladB.

Splicing inhibition triggers global premature termination.

Inhibition of U1 small nuclear ribonucleoprotein (snRNP) triggers strong negative feedback on transcription caused by premature termination *via* the cleavage and polyadenylation machinery (Kaida et al., 2010). We therefore investigated if a similar effect was observed by annotating putative PladB-specific alternative polyadenylation (APA) sites in RNA-seq data. The *BRD4* gene, which regulates the release of

promoter-paused RNA polymerase complexes, showed strong APA up-regulation upon PladB treatment (Figure 3.5A). In mock-treated cells, transcription led to the expression of two long *BRD4* isoforms, as expected (Han et al., 2020). PladB treatment led to significant reduction in the longest isoform and the up-regulation of a short annotated isoform ending after exon 11. In addition, a novel 3′-UTR ending at an annotated PAS site after exon 7 was sharply induced by PladB. RT-qPCR experiments confirmed that the abundance of RNAs overlapping this novel putative APA increased 15-fold 4 hours post PladB treatment, while the long *BRD4* isoform was reduced 2-fold (Figure 3.S5A). Both exon 7 and 11 isoforms ended at or immediately adjacent to an annotated polyadenylation site (PAS) (Figure 3.5A). Systematic annotations revealed 5,771 putative APA sites 4 hours after PladB treatment. Half of these events mapped onto previously annotated 3′UTRs, while the other half defined novel putative intronic APAs (Figure 3.S5B). These candidate intronic APAs were in close proximity to poly(A) motifs (Figure 3.S5C), suggesting they correspond to *bona fide* APA sites. Representative examples from the dataset show clear PladB-specific APAs that were validated by RT-qPCR (Figure 3.S5A, D). Overall, PladB-induced APA events mapped onto 3,631 unique genes showing gene ontology enrichment similar to those observed for down-regulated genes including RNA splicing, cell cycle regulation, and the response to DNA damage stimulus (Figure 3.S5E). Overall, PladB-mediated U2 spliceosome inhibition led to premature termination and profound changes in APA patterns, causing widespread isoform switching.

Increased premature termination should lead to reduced nascent transcription downstream. In agreement, 80% of the genes showing premature APA usage (2,869 / 3,631 genes) showed significant losses in nascent transcription (EU-seq), gene expression (RNA-seq), and/or R-loop levels (DRIP-seq). Metagene plots of EU-seq signal centered at the ends of PladB-specific intronic APAs confirmed increased stalling over the APAs followed by decreased nascent transcription immediately downstream, consistent with termination (Figure 3.5B).

To determine if the three gene classes responded differentially to PladB-induced APA, we plotted the ratios of EU-seq signals observed two and four hours after PladB treatment over the EU-seq signal for the same genes under control conditions (Figure 3.5C). RLL genes showed a minor increase in nascent

transcription very early in gene bodies, followed by a steady decline such that the nascent transcription output under PladB treatment fell below that of control cells within the first ~10% of gene bodies. In contrast, unchanged genes showed a strong increase in nascent transcription immediately downstream of the TSS such that the nascent transcription output under PladB only fell below that of control cells after about 70% through gene bodies. DoG genes had a similar behavior except the nascent transcription output under PladB fell below control only towards the very end of these genes. PladB therefore caused a gradual elongation defect in all three gene classes, as visualized by similar negative slopes in EU-seq ratios. However, the large increase in nascent transcription observed early in gene bodies for unchanged and DoG genes partially compensated for decreased elongation and allowed a significant number of RNA Pol II complexes to reach the end of genes. RLL genes, by contrast, suffered dramatic transcription loss early in elongation. To further address if the reduced elongation was linked to premature APA, we annotated the distance from the TSS to PladB-induced intronic APA events in each gene class. APA events were located in close proximity downstream of the TSS for RLL genes (median distance 3.7 kb, Figure 3.5D). In sharp contrast, DoG genes showed a >4-fold greater median distance between their TSS and intronic APAs (median distance 16.5 kb). Unaffected genes showed intermediate distances. Thus, while all gene classes suffered from PladB-induced premature transcription termination, RLL genes appear particularly vulnerable to it because APA events were encountered significantly earlier during elongation.

PladB treatment leads to increased promoter-proximal pausing.

The observation that *BRD4* underwent premature APA suggested that BRD4 protein levels could be reduced as a result of PladB treatment. Western blots and mass spectrometry assays, however, did not reveal any significant change in overall BRD4 levels over short time intervals (2 and 4 hours). Imaging assays instead revealed that the sub-cellular distribution of BRD4 was significantly altered upon splicing inhibition, with clear BRD4 redistribution towards nucleoli (Figure 3.5E). SF3B2, as expected, was retained in enlarged splicing speckles. Nucleolar sequestration may therefore reduce functional BRD4 pools and affect its ability to mediate pause release at RNA Pol II promoters. In addition, we report that cyclin T1

also became localized to nuclear speckles upon PladB treatment (Figure 3.S5F), possibly reducing functional levels of the critical pause release factor, pTEF-b. We therefore examined patterns of nascent transcription in the vicinity of TSSs as a proxy for pausing. DoG genes showed the most pronounced EU-seq signal immediately downstream of the transcription start site (TSS) (Figure 3.5F, right). This peak was sharply reduced over the first 500-1,000 bp of promoter-downstream sequences, suggesting that DoG genes are heavily regulated by pause-release mechanisms and carry high loads of paused and/or rapidly recycled RNA Pols. Traveling ratios (TRs), which is the relative ratio of RNA Pol II occupancy in the promoter-proximal region and the gene body, (Rahl et al., 2010) confirmed that DoG genes showed significantly higher TRs compared to both unchanged and RLL genes under control conditions (Figure 3.S5G). Importantly, PladB treatment caused a further significant increase in DoG gene promoter-proximal pausing. In sharp contrast, RLL genes had the lowest pausing under control conditions and showed the least gains upon PladB treatment (Figure 3.5G, left). Unchanged genes were more similar to DoG genes in that they showed a strong increase in pausing upon PladB treatment. Thus, the three gene classes defined here solely based on their R-loop behaviors showed important differences in their amounts of promoter-proximal paused RNA Pol II complexes and in their responses to splicing inhibition. Overall, the data suggests that RLL genes are uniquely vulnerable to transcription elongation loss due to their sensitivity to premature APA and their lack of a robust promoter store of paused RNA pol II complexes. In addition, we show here that U2 spliceosome inhibition exerts a negative feedback on RNA polymerase pause release mechanisms, *via* the sequestration of BRD4 and cyclin T1 in distinct sub-nuclear compartments.

DoG genes preferentially show distal splicing alterations which may be coupled to readthrough transcription.

At first glance, DoG and unchanged genes responded similarly to PladB; yet readthrough transcription and DoG R-loops were only seen for DoG genes. Since proper transcription termination has been linked to terminal exon splicing (Davidson and West, 2013; Herzel et al., 2017; Rigo and Martinson, 2008, 2009), we asked whether DoG genes showed a distinct pattern of splicing alterations. DoG genes

showed a pronounced shift of skipped exon events towards the end of gene bodies compared to both RLL and unchanged genes (Figure 3.S5H). Such events often affected terminal or near terminal exons. Similar trends were observed for intron retention events. Interestingly, no single exon gene was present in the DoG gene dataset even though 1,247 single exon genes were expressed in K562 cells and 40 were expected to display DoG transcription at random. These results suggest that DoG genes are significantly more predisposed to aberrant terminal splicing events.

We also wondered if DoG genes might display characteristics predisposing them to weaker transcription termination efficiencies compared to unchanged genes. Analysis of PAS motifs did not reveal significant differences between gene classes. However, we noted that DoG genes showed unusual R-loop patterns through their gene bodies under control conditions. As expected from their high loads of promoter-proximal RNA Pols, DoG genes showed significantly higher R-loop levels over promoter regions and early gene bodies (Figure 3.S5I). However, DoG genes showed significantly reduced R-loop levels towards gene ends compared to other genes. Unchanged genes, in particular, showed a clear rise in terminal R-loops, which has been associated with efficient transcription termination (Proudfoot, 2016; Sanz et al., 2016). Overall, the tendency of DoG genes to experience distal splicing alterations upon PladB treatment combined with intrinsically poorer termination mechanisms may predispose DoG genes to readthrough transcription.

PladB treatment triggers mild DNA damage induction that is temporally and spatially uncoupled from DoG R-loop accumulation

PladB treatment triggers a mild DNA damage response which has been linked to elevated R-loop levels (Nguyen et al., 2019). We showed here that elevated R-loop levels are detected over DoG genes two hours post-PladB treatment and that DoG transcripts can be detected as early as 30 minutes (Figures 3.2, 3.3). We therefore determined if DNA damage induction, as measured by γH2AX immunofluorescence staining, was temporally coupled to DoG R-loop formation. No significant γH2AX foci increase was observed four hours after PladB despite abundant evidence for DoG R-loops. A significant increase in

γH2AX foci only occurred 24 hours post-PladB treatment (Figure 3.6 A,B), although it remained mild compared to the positive control Etoposide (ETO). This suggests that the DNA damage response triggered by PladB is temporally uncoupled from DoG R-loop accumulation. This delay could be explained if DoG R-loops only occur outside of S phase such that cells do not experience the transcription-replication conflicts thought to drive R-loop-mediated genome instability. To address this, cells were synchronized (Figure 3.S6A) and treated with PladB while in G1, S and G2. R-loop levels were then assessed at representative loci for both DoG and RLL genes using DRIP-qPCR. This revealed that PladB-mediated DoG R-loops increases and R-loop losses at RLL genes clearly occurred at all phases of the cell cycle, including S phase (Figure 3.6C). Thus, the lack of induction of DNA damage markers 4 hours post PladB treatment is not due to the lack of DoG R-loops in S phase.

We next asked if the loci experiencing γH2AX deposition at 24 hours would spatially overlap with DoG R-loops, as expected if such R-loops initiate DNA damage. For this, we performed chromatin immunoprecipitation (ChIP) experiments against γH2AX over multiple DoG R-loop hotspots. The DIvA (DSB inducible via AsiSI)/4-hydroxy tamoxifen (4OHT) system developed in U2OS human cells was used to induce double strand breaks (DSBs) at predictable genomic loci (Iacovoni et al., 2010) as a positive control. ChIP-qPCR showed a 25-fold increased γH2AX enrichment in 4OHT-treated cells compared untreated cells (Figure 3.6D). By contrast, γH2AX was not increased over any locus four hours after PladB treatment. After 24 hours, while a trend towards increased γH2AX was observed, this trend was often not statistically significant and was not unique to DoG R-loop loci. Indeed, some RLL regions also showed a trend towards increased γH2AX (Figure 3.6D). Similar results were observed after cells were synchronized along the phases of the cell cycle. Taken together, these results suggest that, although PladB treatment induces DNA damage, such damage does not spatially or temporally correlate with regions that accumulate excessive R-loops.

3.4 Discussion

Intron retention *per se* does not drive R-loop formation

Pharmacological inhibition of the U2 spliceosome using PladB caused the retention of over 7,000 introns in mRNA pools (Figure 3.1), a likely underestimate of the impact of SF3B1 inhibition on splicing at the nascent RNA level (Drexler et al., 2020; Nojima et al., 2015). Despite this, relatively few events of R-loop gains were identified genome-wide and less than 2 percent of annotated retained introns overlapped with R-loop gains. We therefore conclude that inhibiting the U2 spliceosome does not trigger R-loop gains over retained introns. Given that PladB still allows binding of the spliceosome with its target pre-mRNAs (Effenberger et al., 2017; Finci et al., 2018), it is possible that RNA engagement is sufficient to prevent R-loop formation regardless of splicing activity, as was proposed earlier in yeast (Bonnet et al., 2017). However, given that pre-mRNAs are routinely engaged by a plethora of RNA-binding proteins involved in co-transcriptional processing, packaging, and export (Metkar et al., 2018; Singh et al., 2015; Van Nostrand et al., 2020), it is surprising that R-loop formation occurs at all.

We suggest that R-loop initiation occurs over a short window located immediately upstream of the advancing RNA polymerase, prior to protein-mediated RNA binding. The nascent RNA, once it exits the RNA polymerase, is proposed to sample the energy landscape (RNA-DNA base-pairing and DNA topology) such that initiation will occur only if the formation of a short nascent R-loop can withstand the energetic cost of forming a junction between B DNA and a three-stranded R-loop (Chedin and Benham, 2020). Once an RNA sequence has passed this initiation window, we suggest that its probability of interacting with the DNA template decreases sharply, while concomitantly, its probability of being bound by proteins or folded into a secondary structure, increases, further limiting R-loop formation. This co-transcriptional R-loop formation model is consistent with most known aspects of R-loop distributions, including that R-loops most often form co-transcriptionally *in cis*.

SF3B1-mediated splicing inhibition triggers premature termination and promoter-proximal RNA Pol II pausing.

In addition to the expected splicing changes, PladB treatment caused a dramatic loss of transcription capacity affecting over 5,000 genes. We show here that this results from the combined impact of at least two distinct mechanisms. Increased promoter-proximal RNA Pol pausing (Figure 3.5), also observed independently (Caizzi et al., 2021), likely accounts for one such mechanism. Our observation that key pause-release factors, including BRD4 and cyclin T1, become sequestered into distinct nuclear sub-compartments away from transcriptionally active centers, provides a plausible explanation for this effect (Figure 3.7). The relocalization of cyclin T1, a key component of pTEF-b, to splicing speckles is consistent with its less efficient recruitment to promoters (Caizzi et al., 2021). The relocalization of BRD4 to nucleoli is also expected to reduce its functional pools, further lowering pTEF-b recruitment and enhancing promoter-proximal pausing. Importantly, this effect does not appear specific to PladB as a similar accumulation of promoter-proximal transcripts was detected upon treatment of human HeLa cells with Isoginkgetin (IsoG) (Boswell et al., 2017), a non-SF3B splicing inhibitor that targets the U2 spliceosome at a later stage (O'Brien et al., 2008). Overall, increased promoter-proximal pausing due to U2 snRNP inhibition is likely to account for a significant reduction in gene expression given that the release of paused RNA Pol complexes is often a rate-limiting step (Adelman and Lis, 2012; Core and Adelman, 2019; Mayer et al., 2017). Further, prolonged pausing can lower the release frequency and cause an overall reduction in RNA synthesis rates (Gressel et al., 2019; Gressel et al., 2017).

In addition, we show that U2 snRNP inhibition activates global APA, leading to premature termination of these RNA Pol complexes that managed to successfully escape promoter pausing. This second mechanism was evidenced by the steady decline of nascent transcription through gene bodies and by the upregulation of thousands of annotated and novel alternative 3′-UTRs (Figure 3.4). Prior observations that PladB and other SF3B-targeted macrolides reduced the levels of elongation-associated Serine 2 phosphorylated RNA Pol and caused the dissociation of RNA Pol from chromatin (Koga et al., 2015) are consistent with our findings and suggest that these effects are again not specific to PladB. Indeed,

treatment of human cells with IsoG, PladB, as well as other SF3B-targeted splicing inhibitors spliceostatin A and E7107, in addition to U2 snRNP depletion, were shown to cause reduced nascent transcription (Boswell et al., 2017; Caizzi et al., 2021; Han et al., 2022) and induction of APA (Sousa-Luis et al., 2021). Interestingly, inhibition of the U1 snRNP complex also caused dramatic transcription elongation loss associated with premature termination (Berg et al., 2012; Kaida et al., 2010; Oh et al., 2017). Our results and that of others show that SF3B-targeted inhibition of the U2 snRNP complex trigger similar effects, confirming the tight integration of gene expression programs with the splicing machinery through diverse feedback mechanisms (Bentley, 2014; Herzel et al., 2017; Hsin and Manley, 2012).

While the effect of PladB on transcriptional dynamics appears to apply uniformly to all genes, the resulting impact on gene expression outputs varied depending on the intrinsic characteristics of the gene classes identified here (Figure 3.7). RLL genes suffered dramatic early elongation losses which dominated the PladB response landscape. The sensitivity of RLL genes to PladB can be traced to two distinguishing characteristics. First, RLL genes showed the lowest promoter-proximal RNA Pol store and the least increased pausing upon PladB treatment. Thus, any increased termination occurring during elongation cannot be readily compensated by the release of paused RNA Pol complexes. Second, RLL genes carried cryptic PAS motifs shortly downstream of the TSS (median 3.7 kb) compared to other gene classes, ensuring early premature termination. In sharp contrast, DoG genes possessed large loads of paused RNA Pol even under control conditions enabling them to maintain sufficient elongation capacity to complete their transcription. In addition, cryptic PAS sequences only occurred far downstream of the TSS (median 16.5 kb), ensuring that elongation was not interrupted prematurely. Unaffected genes, by virtue of their short size and strong gains of RNA Pol upon PladB treatment, also largely escaped elongation failure in response to PladB. Overall, the dramatic reduction of nascent transcription capacity stemming from RLL genes accounts for the vast majority of co-transcriptional R-loop losses observed upon PladB treatment. This reinforces the notion that R-loop formation is a sensitive readout of nascent transcription.

PladB treatment causes readthrough transcription accompanied by *de novo* R-loop formation.

One additional transcriptional effect of PladB treatment is the induction of readthrough transcription for a subset of 429 DoG genes (Figures 3.2, 3.3). Mechanistically, readthrough transcription may result from splicing alterations, consistent with the links between terminal exon splicing and termination (Davidson and West, 2013; Kyburz et al., 2006; Reimer et al., 2021; Rigo and Martinson, 2008, 2009). Readthrough may also result from a reduction of functional BRD4 pools given recent evidence that BRD4 coordinates the recruitment of 3′-RNA processing factors (Arnold et al., 2021). The characteristics of the DoG gene subclass are consistent with both possibilities. First, DoG genes show a propensity to undergo distal splicing alterations compared to unaffected genes, which may impact the mechanistic coupling between splicing and termination for this gene class. In addition, DoG genes form only low levels of termination-associated R-loops compared to unaffected genes which may predispose them to lower termination efficiencies. Second, DoG genes, with their large stores of promoter-proximal RNA Pols, are heavily regulated via pause release mechanisms and may therefore be most sensitive to a reduction in functional BRD4 levels. Importantly, readthrough transcription was also reported for HeLa cells treated with IsoG (Boswell et al., 2017), suggesting that it is not unique to PladB treatment and may represent a more universal response to U2 snRNP inhibition.

Almost all events of PladB-induced R-loop gains observed mapped to DoG regions, delineating a novel class of aberrant *de novo* R-loops associated with readthrough transcription. Based on current understanding, DoG R-loops are likely to be similar in nature from other co-transcriptional R-loops, with the appearance of long DoG R-loops zones (Figure 3.2) due to the collective piling up of smaller sub-kilobase structures over R-loop prone regions (Malig et al., 2020a). We note, however, that R-loop loads relative to nascent transcription signals, appear significantly higher for DoG regions than for gene bodies. This suggests that DoG transcription may be more prone to R-loop formation, or alternatively, that DoG R-loops are longer lived. Collectively, excessive DoG R-loops occupied nearly 3 megabases of genomic space although, overall, R-loop losses overwhelmingly dominated over R-loop gains. These findings contrast with previous studies reporting that PladB and other splicing inhibitors caused broad R-loops gains (Nguyen et

al., 2018; Nguyen et al., 2017; Wan et al., 2015). Such claims, which relied on S9.6-based imaging approaches, may be problematic given the well-documented issues with S9.6 imaging (Crossley et al., 2019b; Hartono et al., 2018; Phillips et al., 2013; Smolka et al., 2021a). Overall, DoG R-loop gains occurred against the backdrop of a dramatic R-loop reduction genome-wide.

PladB-induced transcriptional alterations resemble other stress responses.

DoG transcription has been observed in response to a range of cellular stresses including oxidative (Giannakakis et al., 2015; Vilborg et al., 2017), heat shock (Cardiello et al., 2018; Vilborg et al., 2017), osmotic (Rosa-Mercado et al., 2021; Vilborg et al., 2015; Vilborg et al., 2017), and viral infections (Bauer et al., 2018; Heinz et al., 2018; Hennig et al., 2018). When compared across stresses, DoG genes display both overlapping and specific gene ontologies that define the cellular response to a given stress (Vilborg et al., 2017). PladB-induced DoG genes were enriched for GOs related to protein translation, response to cellular stress, and mitochondrial energy homeostasis (Figure 3.3). Up-regulated genes identified by RNA-seq showed similar ontologies suggesting that the cellular response to PladB is characterized by attempts at increasing translation capacity, mitochondrial respiration, and ATP synthesis in the face of dwindling and misspliced RNA transcript pools. A second emerging feature of several stress responses is that readthrough transcription occurs in the context of widespread transcriptional repression. Heat shock, for instance, caused decreased RNA Pol II occupancy across most protein-coding genes (Cardiello et al., 2018), accumulation of promoter-proximal paused RNA Pol, and broad gene down-regulation (Gressel et al., 2019; Mahat et al., 2016; Vihervaara et al., 2018). Similarly, osmotic stress led to reduced nascent transcription for thousands of genes (Rosa-Mercado et al., 2021). As described above, PladB treatment also causes dramatic transcriptional shutdown due to increased promoter-proximal pausing and premature termination. Thus, SF3B-targeted splicing inhibition represents a novel example of a cellular stress that triggers increased expression of stress-responsive genes accompanied by readthrough transcription in the face of a global transcriptional shutdown. Interestingly, heat shock is known to trigger splicing inhibition from fly to mammalian cells (Shalgi et al., 2014; Yost and Lindquist, 1986). SF3B1 itself is heat shock-sensitive

and regulates the heat shock response (Kim Guisbert and Guisbert, 2017). This suggests that the heat shock response may, in part, reflect a cellular response to splicing inhibition.

One important aspect of the PladB stress response is the induction of alternative polyadenylation, leading to the increased usage of thousands of novel 3′-UTRs (Figure 3.5). This phenomenon may result from the functional coupling between splicing and 3′-RNA processing, whereby splicing loss unmasks cryptic PAS sites and provides access to the termination machinery. Inhibition of the PRMT1 arginine methyltransferase, which interacts with numerous RNA processing and splicing factors, also led to splicing disruptions and global APA induction (Giuliani et al., 2021). Recent evidence also indicates that the transcriptional shutdown observed under heat shock involves dramatic premature termination (Cugusi et al., 2022). Interestingly, activation of APA is expected to lead to increased isoform switching both at the transcriptome and proteome levels. Whether such changes represent a maladaptive consequence of acute splicing inhibition or play a role in an adaptive stress response (Berkovits and Mayr, 2015; Mayr, 2019; Williamson et al., 2017) remains to be explored. Our observation that BRD4 and cyclin T1 are rapidly relocalized to nucleoli and nuclear speckles, respectively, indicates that protein trafficking may represent another dimension of stress responses. Evidence from heat stress suggests that epigenetic regulators and BRD proteins undergo rapid nucleolar localization (Azkanaz et al., 2019).

Rethinking the links between splicing inhibition, genome instability, and aberrant R-loops.

PladB treatment triggers genome instability as measured by the induction of gH2AX foci (Nguyen et al., 2017; Wan et al., 2015). Mechanistically, the lack of overlap between retained introns and R-loop gains excludes the commonly invoked model that excessive R-loops formed over retained introns underly the PladB genome instability phenotypes. The identification of a novel class of DoG R-loops, however, enabled us to test if these these R-loops formed *de novo* as a result of PladB treatment were driving genome instability. Under our conditions, the DNA damage response was only detectable after 24 hours, while DoG transcription was observed as early as 30 min (Figures 3.2, 3.6). The temporal disconnect between DoG R-loops and DNA damage markers was not due to cell cycle arrest or to DoG R-loops being cell cycle

restricted; DoG R-loops were observed in G1, S, and G2 phases. Considering that toxic R-loops might nonetheless require time to be processed into DNA breaks, we next tested if the gH2AX DNA damage marker, would at least show enrichment over regions that accumulate DoG R-loops. However, this was not the case. Instead, modest gH2AX gains were observed at some, but not all, DoG sites tested and over some RLL genes showing clear R-loop losses. The DNA damage response observed upon PladB treatment was therefore temporally and spatially uncoupled from the accumulation of the common gH2AX marker, raising questions as to whether excessive R-loops represent toxic genomic intermediates under splicing inhibition conditions.

Several non-exclusive alternative models can account for our observations. Transcriptional losses driven by PladB could for instance lead to progressive down-regulation of DDR genes (Figure 3.3), sensitizing cells to spontaneous genomic insults and ultimately resulting in DNA breaks. A similar mechanism was invoked in the context of the response to PRMT1 inhibition (Giuliani et al., 2021) and to the E7107 splicing inhibitor (Han et al., 2022). Conceptually, this predicts that genome instability events may be temporally and spatially uncoupled from R-loop events, as observed here. Alternatively, it is possible that co-transcriptional R-loops as measured by DRIP assays do not represent toxic structures and that instead, another class of R-loops formed at paused promoters drive the genomic conflicts from which DNA damage arises (Castillo-Guzman and Chedin, 2021). In addition, we cannot rule out that the increased R-loop loads driven here upon SF3B1 inhibition were not sufficient to mediate an effective R-loop-mediated DNA damage response. Analysis of additional cellular models of R-loop dysfunction will address this issue.

Overall, our findings put into question the model that proper splicing is necessary to protect the genome from the damaging consequences of excessive genic R-loops (Li and Manley, 2005; Paulsen et al., 2009; Tresini et al., 2015). Our study also highlights the necessity of directly characterizing R-loop patterns at the genome scale and of factoring the dynamic cellular and transcriptional responses that likely accompany any genetic or pharmacological perturbation targeted to RNA processing factors.

3.5 Figures
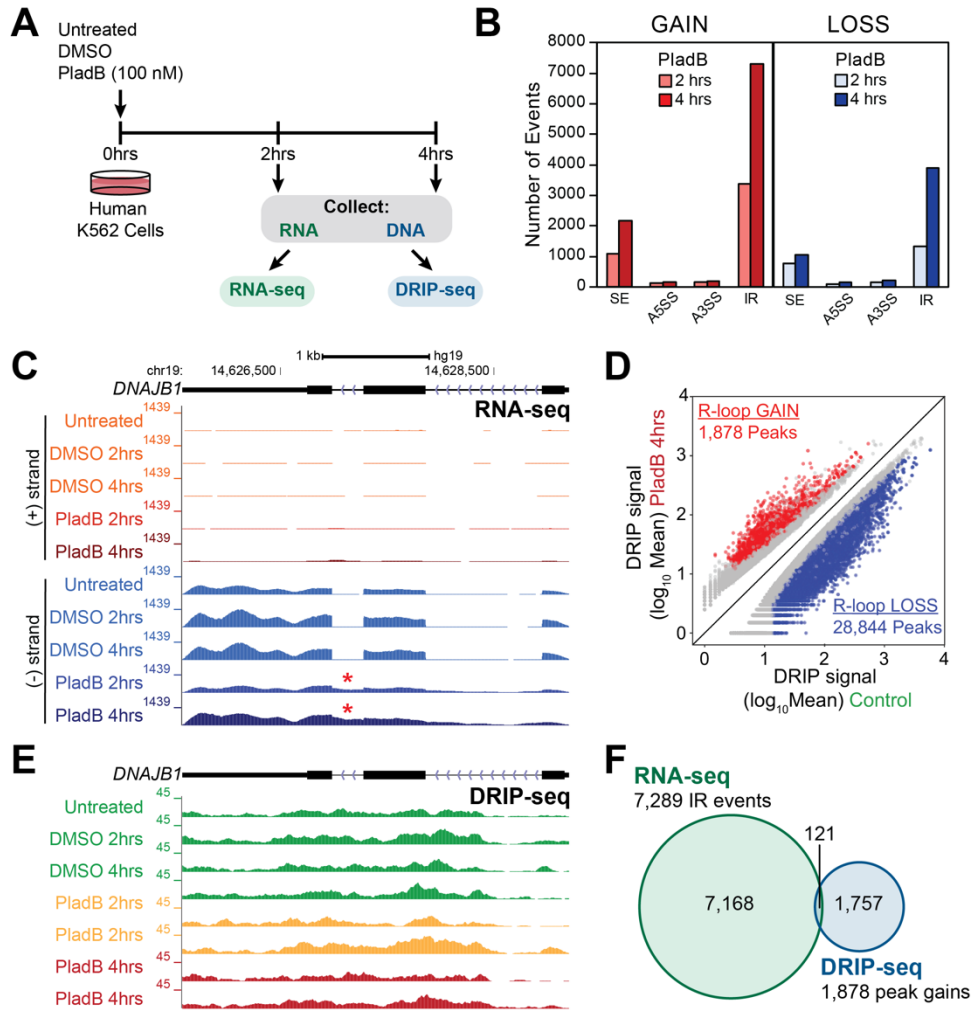
Figure 3.1: A. Schematic of experimental setup, see text for details. B. Summary of PladB-mediated splicing alterations identified by RNA-seq. The number of skipped exons (SE), alternative 5' splice sites (A5SS), alternative 3´ splice sites (A3SS) and intron retention (IR) gained or lost after 2 and 4 hours of PladB treatment are graphed. C. Genome browser screenshot over the representative *DNAJB1* gene showing plus and minus strand poly(A) RNA-seq signal obtained from controls and PladB treated K562 cells. Intron retention is observed two- and four-hours post PladB treatment (asterisk). D. XY plot of normalized DRIP-seq counts for control and treated (PladB 4 hours) samples. Each dot on the graph corresponds to a DRIP-seq peak. Red/blue colors indicate significant gains/losses of signals (adjusted p-value of < 0.05 and

$|\log_2(\text{FC})| \geq 1$). E. Genome browser screenshot showing the same region as in (C) but now displaying DRIP-seq signals. No increase in R-loop signal is detected over the retained intron. F. Venn diagrams displaying the intersection of events of intron retention and R-loop gains.
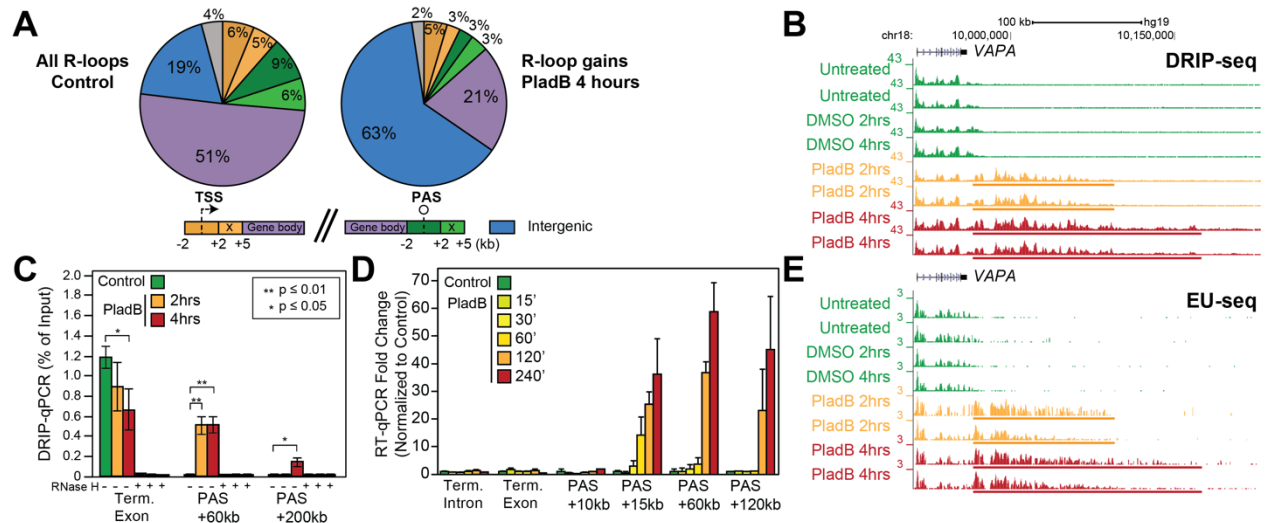
Figure 3.2: A. Location analysis of DRIP peak gains for control (left) and PladB-treated samples (right) across genomic compartments (color-coded according to the schematic below). R-loop distribution across compartments is indicated by percentages. TSS, transcription start site; PAS, polyadenylation site. B. Genome browser screenshot over the representative *VAPA* gene and ~200 kb downstream showing DRIP-seq signal from controls and PladB-treated K562 samples. R-loop gains occur directly downstream of gene (DoG - colored bar). C. Bar chart of DRIP-qPCR (as percent of input) for *VAPA* and two regions downstream of PAS. Each bar is the average of three-independent experiments (shown with SE). D. Bar chart of RT-qPCR (as fold change over control) at indicated loci through a time course after PladB treatment. Each bar represents the average of three independent experiments (shown with SD). E. Genome browser screenshot over the same region as in (B) but now displaying EU-seq signals. Regions with increased R-loops also have increased nascent transcription.
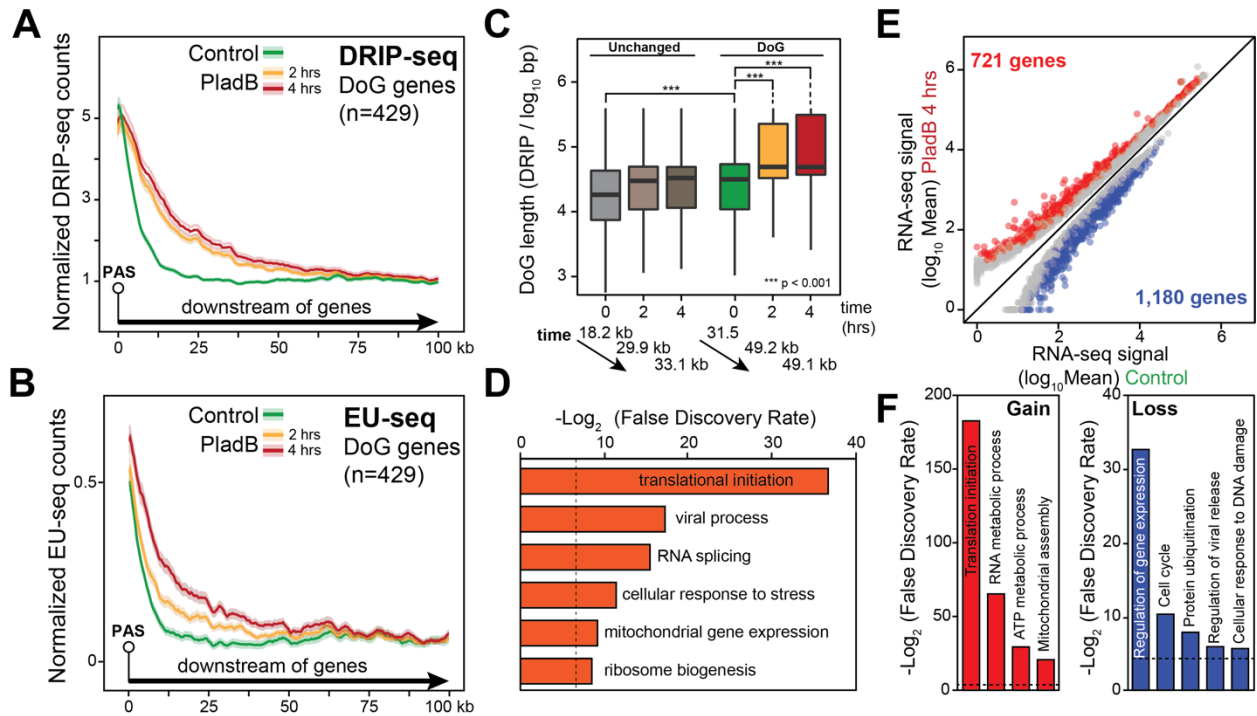
Figure 3.3: A,B. Metaplots of DRIP-seq (top) and EU-seq (bottom) signals extending from nearest PAS for all annotated DoG genes. For each time point, the signal is shown as a trimmed mean (line) surrounded by SE (shaded). C. DoG R-loop lengths (measured from DRIP-seq data) under control conditions and two- and four-hours post PladB for DoG genes and expression-matched genes that showed no significant R-loop changes. P-values were determined by a Wilcoxon Mann-Whitney test. The numbers below indicate median terminal R-loop lengths at each time points. D. Enriched gene ontologies for DoG genes. The dashed line indicates 1 % FDR. E. XY plot of normalized RNA-seq counts for control and treated (4 hours PladB) samples. Each dot on the graph corresponds to a gene. Red/blue colors indicate gains/lo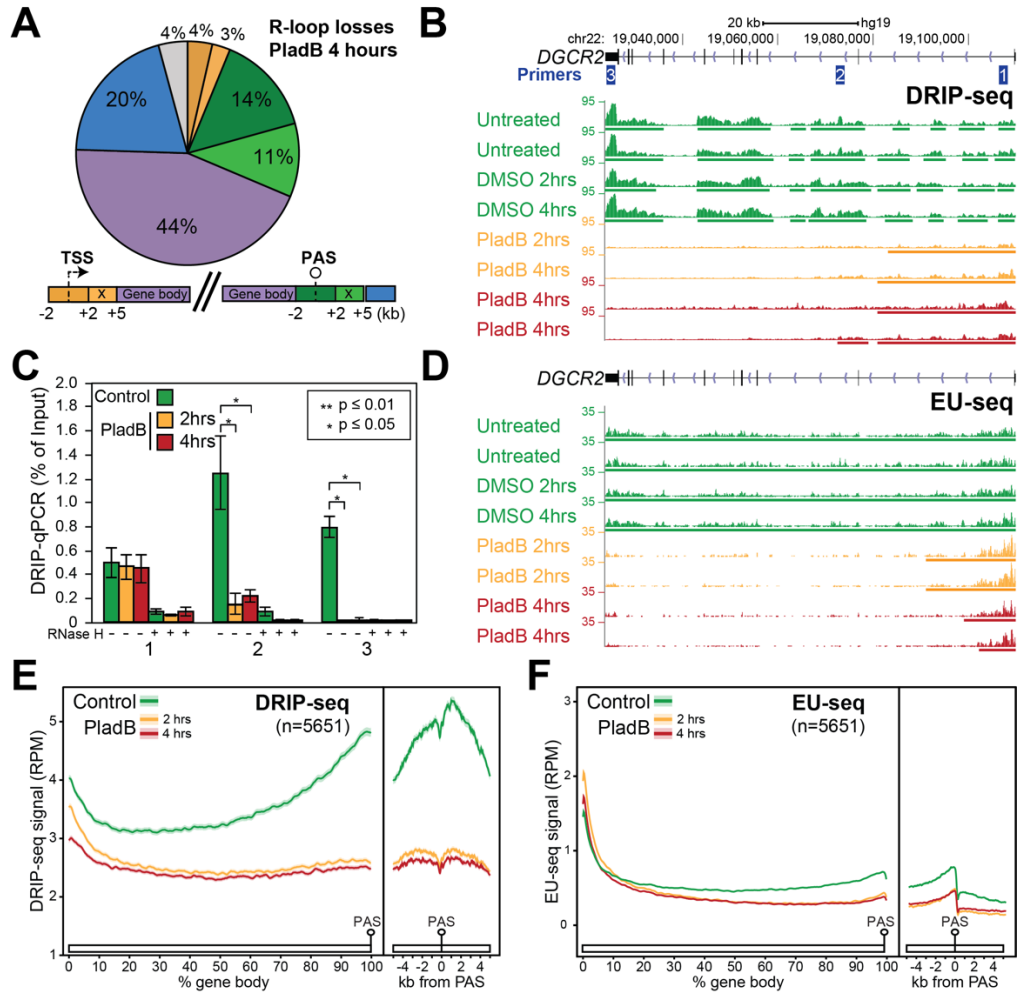sses of RNA-seq signals (adjusted p-value of ≤ 0.05 and $|\log_2(FC)| \geq 1$). F. Enriched gene ontologies for RNA-seq gain genes and for RNA-seq loss genes. The dashed line indicates 5% FDR.

Figure 3.4: A. Location analysis of R-loop losses for PladB-treated (4 hours) samples across genomic compartments. Numbers indicate the percentage occupied by each compartment. B. Genome browser screenshot over the representative *DGCR2* gene showing DRIP-seq signal from controls and PladB-treated K562 samples (colored bars indicate R-loop peaks). C. Bar chart of DRIP-qPCR (as percent of input) for *DGCR2* using indicated qPCR primer pairs. Each bar is the average of three-independent experiments (shown with SE). D. Genome browser screenshot showing the same region as in (B) but now displaying EU-seq signals. E. Metaplot of DRIP-seq signals over gene body (as % of gene length) and terminal regions (+/- 5kb from PAS) of RLL genes. Control and PladB-treated samples are color-coded as indicated. For each time point, the signal is shown as a trimmed mean (line) surrounded by the standard error (shaded). F. Same as E except EU-seq signals are plotted.
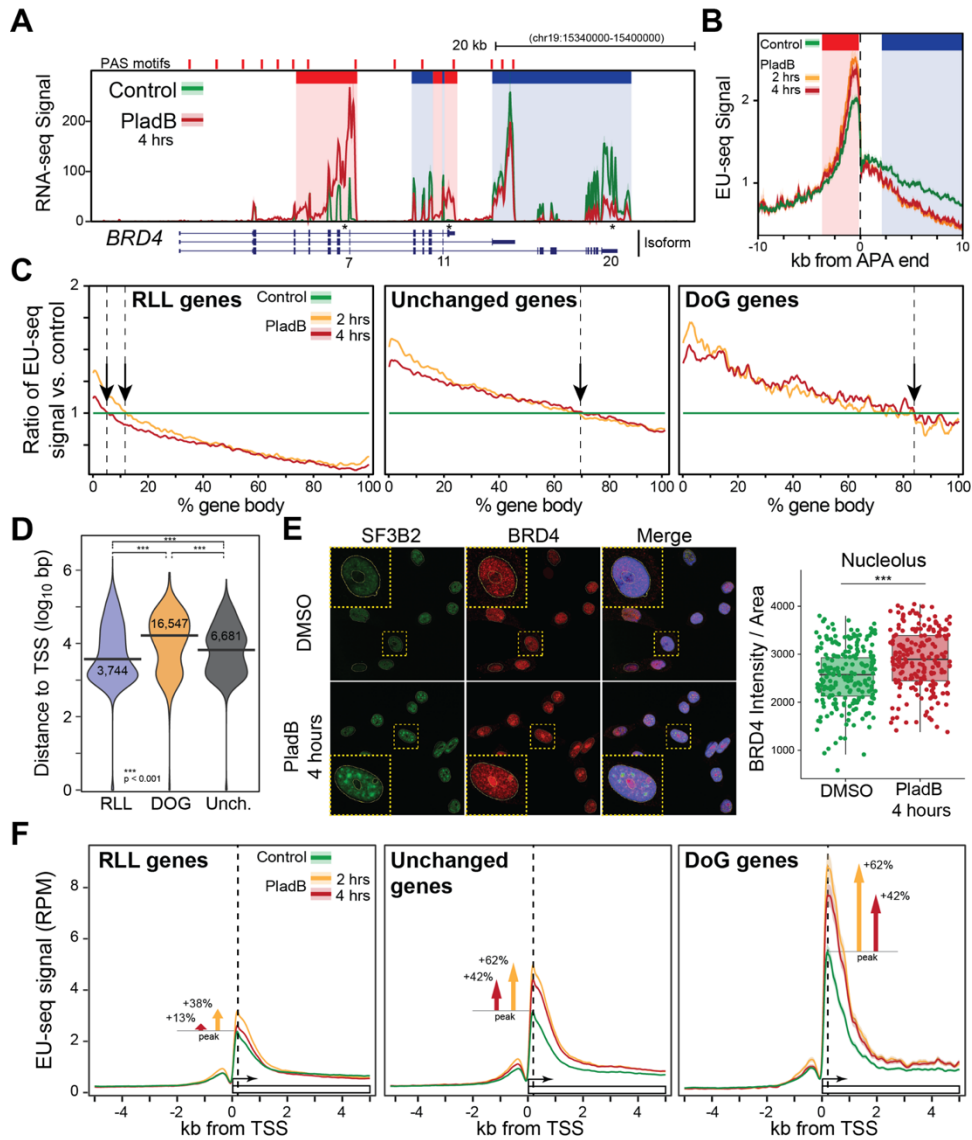
Figure 3.5: A. RNA-seq signal for *BRD4* under control and PladB-treated conditions. Highlighted in red and blue are sights of significantly increased or decreased APA usage (adj. p-value < 0.05). B. Metaplot of EU-seq signals over APA end of all genes (+/- from APA end) under control and PladB-treated conditions. C. Ratio of mean EU-seq signals (PladB/control) for DoG, unaffected, and RLL genes plotted over gene body for control and PladB-treated conditions. The positions at which PladB nascent transcription signals fall below control signals are highlighted by vertical arrows. D. Distance from the TSS to PladB-induced intronic APA events between DoG, unaffected, and RLL genes. E. Representative immunofluorescence

microscopy images from DMSO- and PladB-treated (4 hours) analyzed for SF3B2 and BRD4 distribution and DAPI. Quantification of the immunofluorescence data for each condition; the number of cells across two biological replicates analyzed per condition is indicated. F. Metaplot of EU-seq signals over DoG, unaffected, and RLL gene promoters (+/- 5 kb from TSS) under control and PladB-treated conditions. The relative signal increases measured at the promoter-proximal peak and 2 kb downstream of TSS are indicated.
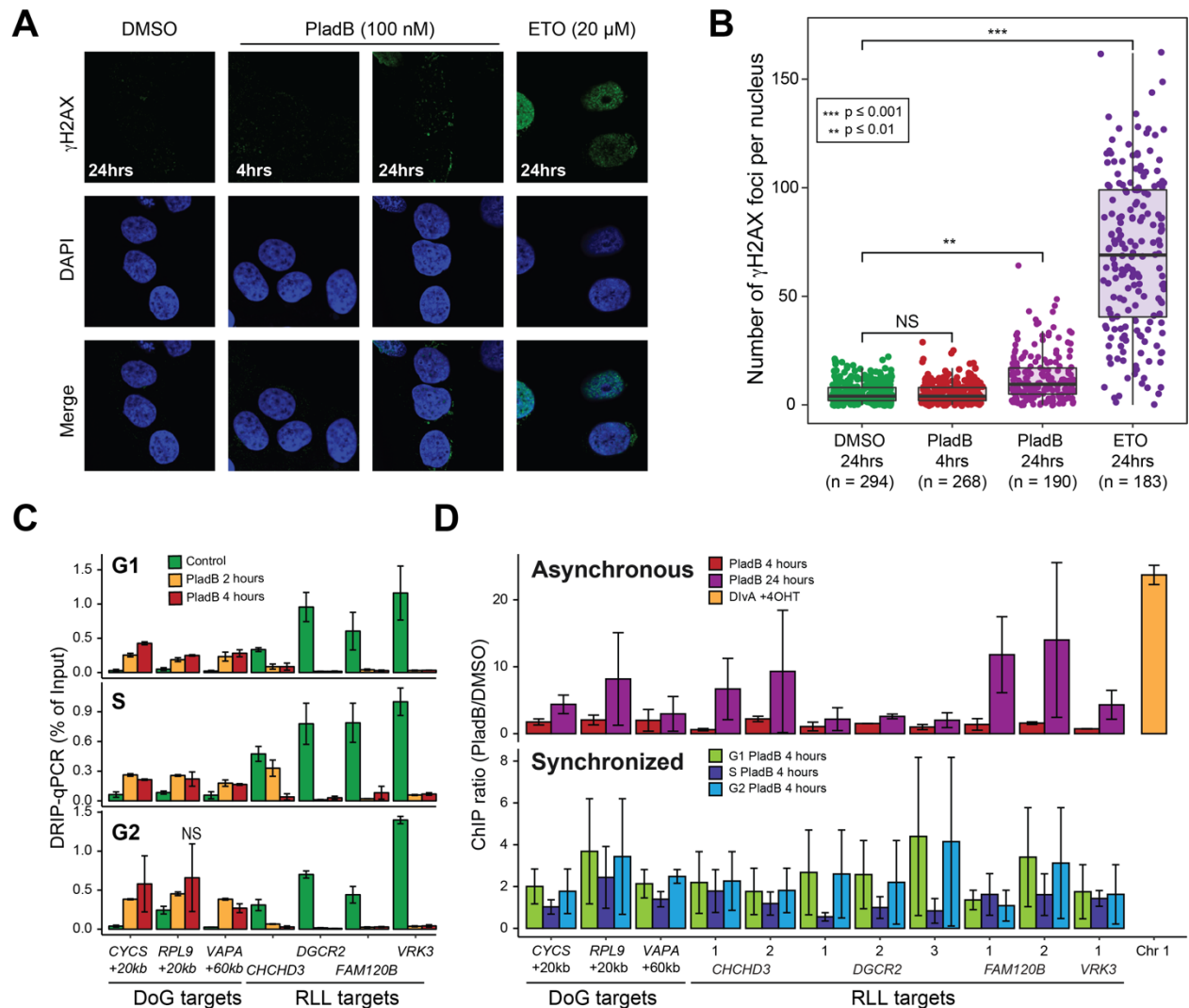
Figure 3.6: A. Representative immunofluorescence microscopy images from DMSO- and PladB-treated (4 and 24 hours), and Etoposide-treated (ETO) HeLa cells analyzed for γH2AX distribution and DAPI. B. Quantification of the immunofluorescence data for each condition; the number of cells across two biological

replicates analyzed per condition is indicated. C. Bar chart of DRIP-qPCR (as percent of input) for DoG genes and RLL genes for synchronized cells in G1, S, or G2/M. Each bar is the average of two-independent experiments (shown with SE). D. γH2AX ChIP-qPCR data expressed as ratio (PladB over DMSO for γH2AX) over a range of loci, treated or not with PladB and for cells synchronized in G1, S or G2/M, as indicated. IgG was used as a non-specific negative control on the same samples. The DIvA cell line (shown at right) in which site-specific breaks can be induced was used as a positive control for the experiment. The induction of γH2AX recruitment in DIvA cells was statistically significant (p<0.05). All other changes were not statistically significant.
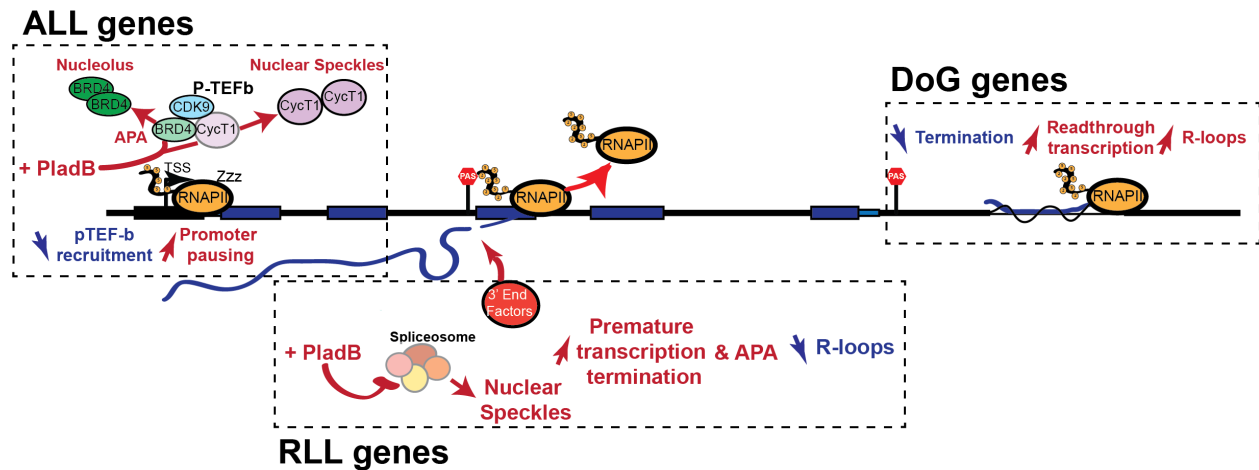
Figure 3.7. Model. PladB treatment triggers reduction in pTEF-b recruitment and increase in promoter-proximal pausing for all genes, driven in part by the relocalization of Cyclin T1 and BRD4. PladB treatment triggers increase in premature transcription termination and APA, resulting in a global decrease in R-loops at RLL genes. PladB also triggers decrease in termination, and readthrough transcription resulting in the formation of DoG R-loops for a subset of genes. See text for details.

3.6 Methods

Analysis of transcriptome alterations by RNA-seq and RT-qPCR

Human K562 cells were treated with PladB (100 nM final concentration) for 2 and 4 hours, at which point they were harvested. Mock-treated (DMSO) and untreated samples were also collected as controls. At each time point, total RNA was isolated from 1×10^6 cells using TRI reagent (Life Technologies, Grand Island, NY) and Direct-zol RNA MiniPrep kit (Zymo Research). Isolated RNA was then treated with 2.5 µl of DNase I (NEB) for 1 hour at 37°C. Following DNase I treatment, RNA was cleaned up using RNA Clean & Concentrator-5 Kit (Zymo Research) and resuspended in 15 µl. After cleanup, poly(A) RNA was selected and RNA-seq libraries were constructed using the Illumina TruSeq kit prior to sequencing on an Illumina HiSeq 4000 instrument. Mapping and all downstream analysis were performed on two independent biological replicates.

For RT-qPCR, cDNA was generated using iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad), with 900 ng of human RNA and 100 ng of Hydra RNA (as exogenous normalizer). cDNA was diluted 1:3 (final volume 60 µl). For RT-qPCR, 1 µl of 1:3 dilution of the cDNA was used per well in a 20 µl reaction, along with 2 µl of 10 µM primer sets, and 10 µl of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). Reactions were run in triplicate on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with the following protocol: 95 °C (30 sec), 40 cycles of 95 °C (10 sec) and 95 °C (30 sec), followed by melt curve analysis (65-95 °C). Quantification was calculated using the $\Delta\Delta C_t$ method. All values were normalized to a Hydra gene (*RP49*) and to time zero (Untreated/DMSO). RT-qPCR was performed on three independent biological replicates.

Analysis of R-loop genomic distribution by DRIP-seq and DRIP-qPCR

PladB treated and control K562 cells were harvested 2 and 4 hours after PladB treatment. At each time point, genomic DNA was isolated from 5×10^6 cells and extracted as described (Sanz and Chedin, 2019b). S9.6-based DNA:RNA hybrid immunoprecipitation (DRIP) was performed as described previously (Ginno et al., 2012). In brief, 4.4 µg of restriction digested genomic DNA was immunoprecipitated for 16 hours at 4°C with 10 µL of S9.6 antibody. Three immunoprecipitated samples were combined into one for library construction. Once the quality of the combined immunoprecipitation was checked by qPCR at a range of positive and negative test loci, barcoded sequencing libraries were built, library quality was verified, and sequencing was performed on Illumina HiSeq or NovaSeq instruments. Mapping and all downstream analysis was performed on two independent biological replicates. R-loops were also mapped using sDRIP-seq, a variant of DRIP-seq that permits strand-specific R-loop mapping after genome fragmentation using sonication (Smolka et al., 2020).

For DRIP-qPCR, 2 µl of immunoprecipitated DNA (and 1:10 diluted input DNA) was used per well in a 20 µl reaction, along with 2 µl of 10 µM primer sets, and 10 µl of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). Reactions were run in duplicate on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with the following protocol: 95 °C (30 sec), 40 cycles of 95 °C (10 sec) and 95 °C (30 sec), followed by melt curve analysis (65-95 °C). All values were normalized relative to input and normalized to Control (Untreated & DMSO) to calculate fold enrichment. DRIP-qPCR was performed on three independent biological replicates.

For RNase H treatment, 6 µg of restriction digested DNA was treated with Rnase H1 for 1 hour at 37 °C using a 1/100 dilution of the stock enzymes in 50 mM Tris-HCl pH 7.5, 75 mM KCl and 3 mM MgCl2.

DNA was then purified with phenol/chloroform and ethanol precipitated before being used in S9.6 immunoprecipitations as described above.

Analysis of nascent transcription using EU-seq

Human K562 cells treated or not with PladB were incubated with 5-ethynyl uridine (EU) (0.5 mM final concentration) for 20 min prior to harvesting two- and four-hours post PladB treatment. At each time point total RNA was isolated from $3x10^6$ cells and initially processed as described for RNA-seq. EU-RNA was enriched using Click-iT Nascent RNA Capture Kit (Invitrogen), with 5 µg of EU-RNA for biotinylation by click reaction and 1 µg of biotinylated RNA for binding to Streptavidin T1 magnetic beads. cDNA was generated using iScript Select cDNA Synthesis Kit (Bio-Rad). cDNA was cleaned up before second-strand synthesis. Then cDNA was ligated to Illumina barcoded adapters for sequencing. Library quality was checked on an Agilent BioAnalyzer and sequencing were performed on Illumina HiSeq or NovaSeq instruments. Mapping and downstream analysis was performed on two independent biological replicates.

Sequencing data processing and computational data analysis.

STAR version 2.7 (2.7.0f_0328) was used to map all sequencing reads and splice junctions (default parameters). Macs2 v2.2.5 was used to call EU-seq and sDRIP-seq signals according to defined bins. For DRIP-seq, we used restriction digest as bins. For RNA-seq, we used annotated genes and gene features (e.g. exon, intron) as their bins using Gencode version 19 (GRCh37). For subsequent analysis, the gene set was further filtered for highest APPRIS principal protein-coding genes with expression levels of at least 10 RPKM, as an annotation file. Htseq-count v0.11.2 was then used to count reads into their bins. DESeq2 v1.24 was used to calculate significant gains or losses (adjusted p-value of ≤ 0.05 and $|log2(FC)| \geq 1$). rMATS v4.0.2 were used to identify significant changes (adjusted p-value of ≤ 0.05 and $|log2(FC)| \geq 0.32$) in skipped exons (SE), A3SS, A5SS, and intron retentions (IR). In addition to rMATS, intron retention

events were called by identifying significant increases in intronic reads using DESeq2 as well as splice junctions from STAR. To distinguish IR events from other possible events, intronic gains that spanned less than 75% of intron length were categorized into putative 3´UTRs if their 5´ end were adjoined (within +/- 10% of intron length) to the 3´ end of an exon or junction, and their 3´ end were not adjoined to the 5´ end of an exon or junction (and vice versa for putative 5´UTRs). For location analysis, we divided the genic space into several regions: promoters (+/-1kb of TSS), gene body (+1kb of TSS to -1kb of PAS), terminal (+/- 1kb of PAS), downstream of genes (DoG), and intergenic regions (everything else). DoG regions were defined as regions between a gene's terminal region (+1kb of PAS + 1 kb) up to either 150 kb downstream if no neighbor was found prior, or to the promoter of its closest downstream gene minus 1 kb, whichever is closest. The 150 kb limit was chosen because a few test genes showed DoG regions extending more than 100 kb. Post-analysis, very few genes had DoG regions longer than 100 kb. To correctly assign a DoG region to the gene from which it originated, we annotated the strandedness of EU-seq and DRIP-seq signals based on the ratio of positive and negative strands of RNA-seq and sDRIP-seq signals.

Fixation and labelling for γH2AX, BRD4, SF3B2 immunofluorescence

HeLa cells were grown, fixed, permeabilized, washed, immunostained, and imaged in 35 mm glass bottom poly-D-lysine-coated dishes (P35GC-1.5-14-C, MatTek) using 2 mL volumes of media and buffer solutions. All steps for fixation and immunofluorescence were carried out at room temperature. Cells were fixed in freshly prepared 1% formaldehyde in PBS for 10 minutes. Fixation solutions were quenched with the addition of 200 µL of 1 M glycine in PBS. Samples were washed once with PBS, and then incubated in permeabilization buffer (PBS with 0.1 % Triton X-100) for 10 minutes. Samples were then incubated in staining buffer (TBST with 0.1 % BSA (A9647-50G, Sigma)) for 10 minutes with rocking and a 1:1,000 dilution of primary anti-phospho-Histone H2A.X (Ser139) antibody (05-636, Sigma) was added for 1 hour with rocking. Samples were then washed once with staining buffer and incubated with a 1:2000 dilution of secondary anti-mouse AlexaFluor 488 conjugate (A28175, Invitrogen) for 1 hour with rocking with samples

kept concealed from light from this step onward. A 2.5 µg/mL DAPI solution in staining buffer was then added for 2 minutes and washed in TBST for 10 minutes with rocking. Samples were then immediately imaged. For each experiment, all samples were prepared, treated, and stained in parallel from one master antibody and/or dye dilution to ensure uniform treatment and staining efficiencies. Fixation and labelling was performed for two independent biological replicates.

Imaging and image analysis

Fixed cell imaging was performed using a confocal laser scanning biological microscope [Olympus FV1000] with a 60X objective. For each experiment, all conditions were imaged in parallel with identical exposure times and laser settings. Images were analyzed and quantified using ImageJ. Statistics and data visualization were done using RStudio. P-values were determined by a Wilcoxon Mann-Whitney test using wilcox.test() function. The number of foci were counted for individual nuclear regions defined by thresholding the DAPI channel to obtain Regions of Interest (ROI) for analysis. ROIs containing more than one nucleus were removed from downstream analysis. Find Maxima was applied to the 488 channel using a noise tolerance of 700 to generate a binary image of local maxima (γH2AX foci). Analysis was then performed on nuclear ROIs to quantify the number of γH2AX foci per nucleus.

Analysis of γH2AX genomic distribution using ChIP-qPCR

Human K562 cells ($1x10^7$ cells) were treated with PladB (100 nM final concentration) for 4 and 24 hours or DMSO as mock-control. DIvA cells were treated with or without 4-hydroxy tamoxifen (4OHT) (300 nM final concentration) for 4 hours. K562 cells were then collected and resuspended in PBS, then formaldehyde was added to a final concentration of 1% and crosslinking was allowed to proceed for 10 minutes at room temperature, while the samples were rocking. For DIvA cells, media was removed and replaced with PBS

and Formaldehyde was added to dish. In order to stop the reaction, fresh glycine was added to a final concentration of 0.125 M. For DIvA cells, the cells were then scraped and washed once with PBS. After 10 minutes cells were washed once with PBS. Pelleted cells were then resuspended in RIPA lysis buffer (150 mM NaCl, 5 mM EDTA, 50 mM Tris pH 8.0, 1% IGEPAL CA-630, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with phosphatase and protease inhibitors (PHOSS-RO, Sigma and 539134, Millipore). The lysate was then sonicated 30 secs for 15 cycles (Diagenode Bioruptor) to obtain DNA fragments of about 500-1,000 bp. Samples were then centrifuged for 15 mins at 4ºC at 15,000 rpm and the supernatants were transferred to new tubes. Samples were then diluted in RIPA buffer (100 µl chromatin : 900 µl RIPA buffer) and phosphatase and protease inhibitors were added to samples. A portion of the sample was removed and saved as Input. The rest of the sample was subjected to a 60 min preclearing with 60 µl of previously blocked protein A/G magnetic agarose beads (PI78609, Thermo Scientific) at 4 ºC. Blocking was achieved by incubating the agarose beads in RIPA buffer with 0.1 % BSA for 30 minutes. Precleared samples were incubated with rotation overnight at 4 ºC with either 10 µg of anti-phospho-Histone H2A.X (Ser139) antibody (05-636, Sigma) or 10 µg of IgG as control. Immune complexes were then recovered by incubating the samples with 60 µl of previously blocked protein A/G magnetic agarose beads for 1 hour at 4ºC on a rotating wheel. Beads were washed 6 times in wash buffer (150 mM NaCl, 10 mM Tris pH 7.5, 2 mM EDTA, 0.1% IGEPAL CA-630) for 10 minutes followed by an additional wash with 10 mM Tris-EDTA pH 8.0. Elution and crosslink reversal were achieved by incubating samples in 10 mM Tris-EDTA + 0.1% SDS and proteinase K at 65ºC for 4 hours. Elutions were then transferred to a phase-lock separation tube. The beads were then washed one more time with 10 mM Tris-EDTA and 0.5 M NaCl for 10 minutes at 65ºC and the supernatant was added to the phase-lock separation tube. DNA was extracted by the phenol-chloroform method and precipitated with ethanol.

For ChIP-qPCR, 2 µl of immunoprecipitated DNA (and 1:4.5 diluted input DNA) was used per well in a 20 µl reaction, along with 2 µl of 10 µM primer sets, and 10 µl of SsoAdvanced Universal SYBR Green

Supermix (Bio-Rad). Reactions were run in duplicate on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with the following protocol: 95 °C (30 sec), 40 cycles of 95 °C (10 sec) and 95 °C (30 sec), followed by melt curve analysis (65-95 °C). All values were normalized relative to input. ChIP-qPCR was performed on two independent biological replicates for K562 samples and two technical replicates for DIvA samples.

Cell cycle synchronization

To obtain G1 stage-specific synchronization, sub confluent K562 cells were treated for 24 hours with thymidine (2 mM) and released into pre-warmed medium for 3 hours before adding nocodazole (100 ng/mL) for 12 hours. Then the cells were released for either 8 or 10 hrs before adding PladB (100 nM) for 4 or 2 hours, respectively. The cells were then collected for both flow cytometry and DRIP-qPCR. Each experiment was performed two times.

To obtain S stage-specific synchronization, sub confluent K562 cells were treated for 24 hours with thymidine (2 mM) and released into pre-warmed medium containing PladB (100 nM) for 2 or 4 hours. The cells were then collected for both flow cytometry and DRIP-qPCR. Each experiment was performed two times.

To obtain G2/M stage-specific synchronization, sub confluent K562 cells were treated for 24 hours with thymidine (2 mM) and released into pre-warmed medium for 6 hours before adding RO-3306 (10 mM) and PladB (100 nM) for 2 and 4 hours. The cells were then collected for both flow cytometry and DRIP-qPCR. Each experiment was performed two times.

Cell Cycle Profiles

For fixation, $2x10^6$ K562 cells were washed twice and resuspended in 1 mL of PBS. This mixture was slowly added to 4 mL of 100 % ethanol. Cells incubated on ice for 30 minutes. The fixed cells were centrifuged at 800 x g for 5 minutes and washed once with PBS. The fixed cells were then stained with 600 ul of Propidium Iodide (PI). The samples were detected with an BD FACS Canto II Flow Cytometer. Cell-cycle distribution was analyzed using the FlowJo Software with a minimum of 10,000 cells. Each experiment was performed two times.

3.7 Supplemental Figures
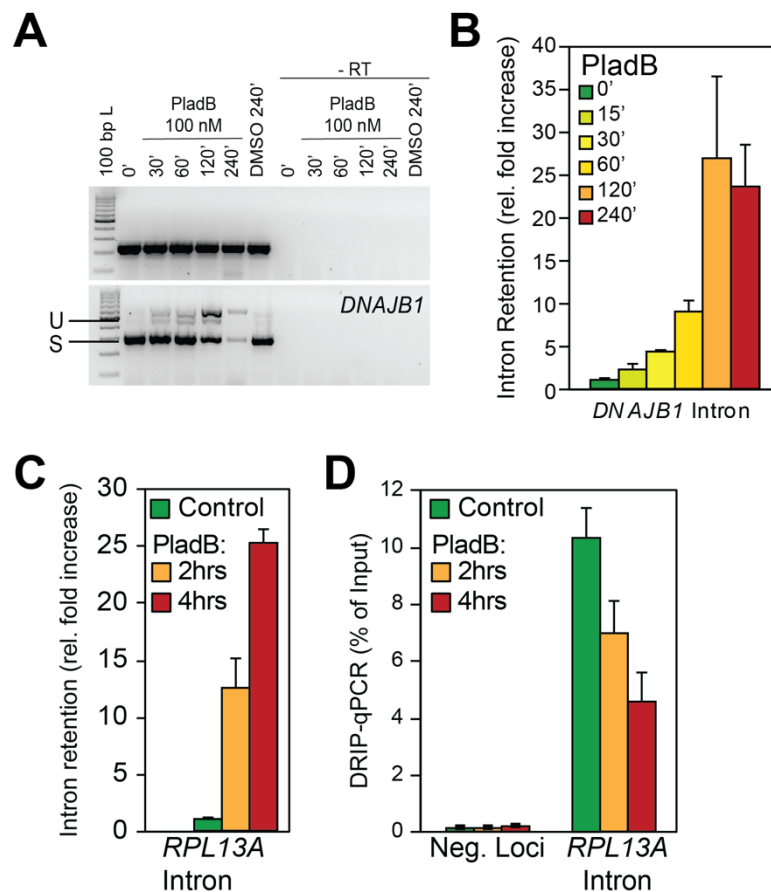
Figure 3.S1: A. Gel-based RT-PCR splicing assay verifies that PladB treatment causes intron retention over the *DNAJB1* target gene, but not the control *ACTB* gene. No PCR products were obtained if reverse-transcription (RT) was omitted, as expected. B. RT-qPCR assay confirms that PladB treatment causes intron

retention for *DNAJB1* target gene. IR starts to occur as early as 15 min. C. RT-qPCR assay confirms that

PladB treatment causes intron retention for *RPL13A*, a positive R-loop forming gene. D. Bar chart of DRIP-

qPCR (as percent input) for *RPL13A*. Each bar is the average of three-independent experiments (shown

with SE). PladB treatment results in loss of R-loops over intron with IR.
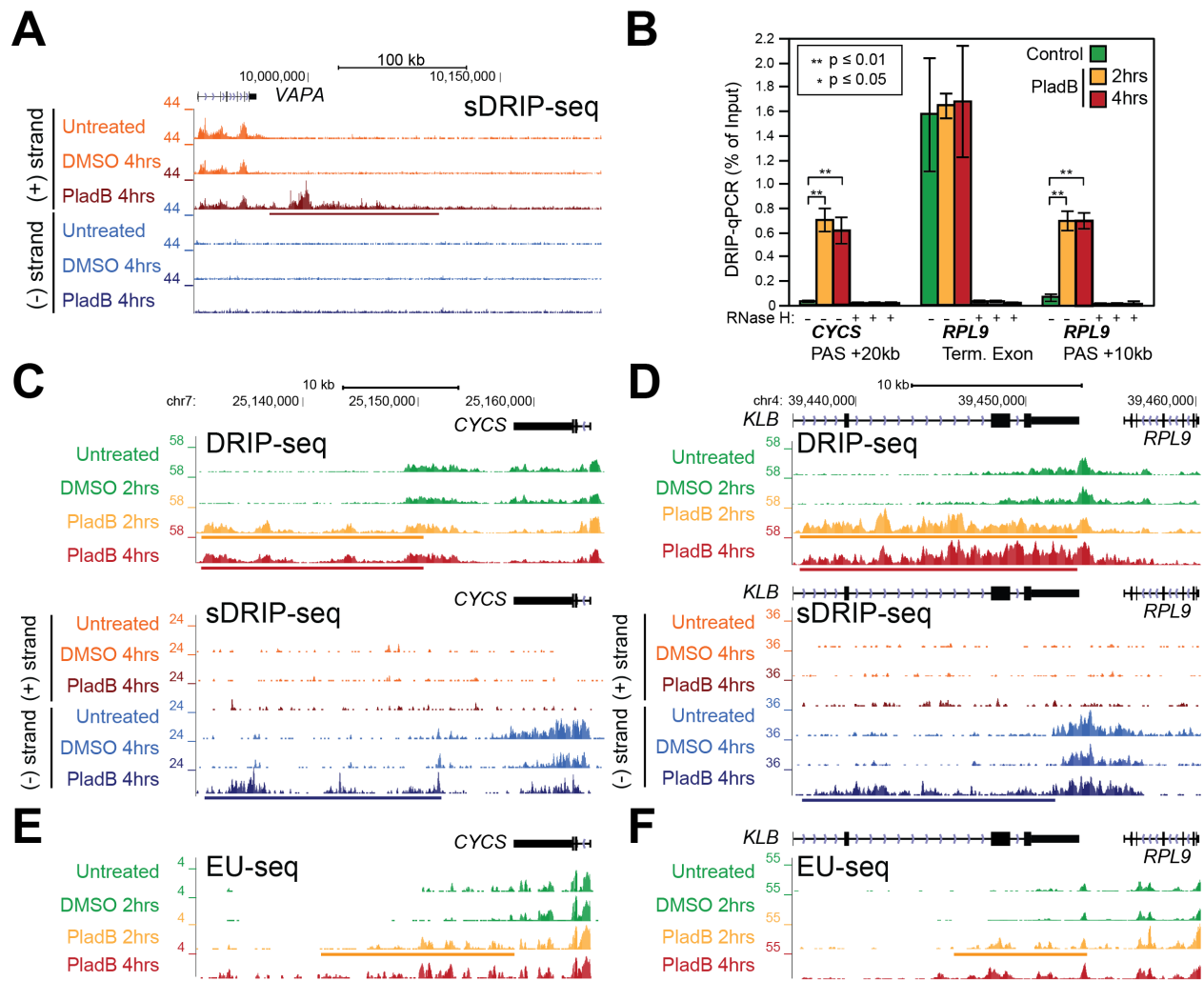
Figure 3.S2: A. Genome browser screenshot over the *VAPA* gene and ~200 kb downstream showing plus

and minus strand sDRIP-seq signal obtained from controls and PladB-treated K562 samples. R-loop gains

occur directly downstream of gene (DoG) only on the plus strand (colored bar). B. Bar chart of DRIP-qPCR

(as percent of input) for for a region downstream of *CYCS* PAS, *RPL9* and a region downstream of *RPL9*

PAS. Each bar is the average of three-independent experiments (shown with SE). C. Genome browser

screenshot over the *CYCS* gene and downstream region showing DRIP-seq signal (Top) and plus and minus strand sDRIP-seq signal (Bottom) obtained from controls and PladB treated K562 samples. R-loop gains occur directly downstream of gene (DoG) and only on minus strand (colored bar). D. Genome browser screenshot over the *RPL9* gene and downstream region showing DRIP-seq signal (Top) and plus and minus strand sDRIP-seq signal (Bottom) obtained from controls and PladB treated K562 samples. R-loop gains occur directly downstream of gene (DoG) and only on minus strand (colored bar). E. Genome browser screenshot showing the same region as in (B) but now displaying EU-seq signals. Region with increased R-loops also have increased nascent transcription. F. Genome browser screenshot showing the same region as in C but now displaying EU-seq signals. Region with increased R-loops also have increased nascent transcription.

Figure 3.S3: A. Location analysis of R-loop peak gains for PladB-treated K562 sample (4 hours) according to downstream of gene (DoG) annotation pipeline. Annotation scheme is below. 82% of R-loop gains 4 hours post-PladB treatment specifically map to downstream of gene (DoG) regions. B. Metaplots of sDRIP-seq signals extending from nearest PAS for all annotated DoG genes. For each sample, the signal is shown as a trimmed mean (line) surrounded by SE (shaded). C. STRING interaction analysis. All 429 DoG genes were used as input for STRING analysis and a network was built based on highest confidence (0.9). Nodes are color coded based on biological processes. Nodes that had no interaction are not shown.
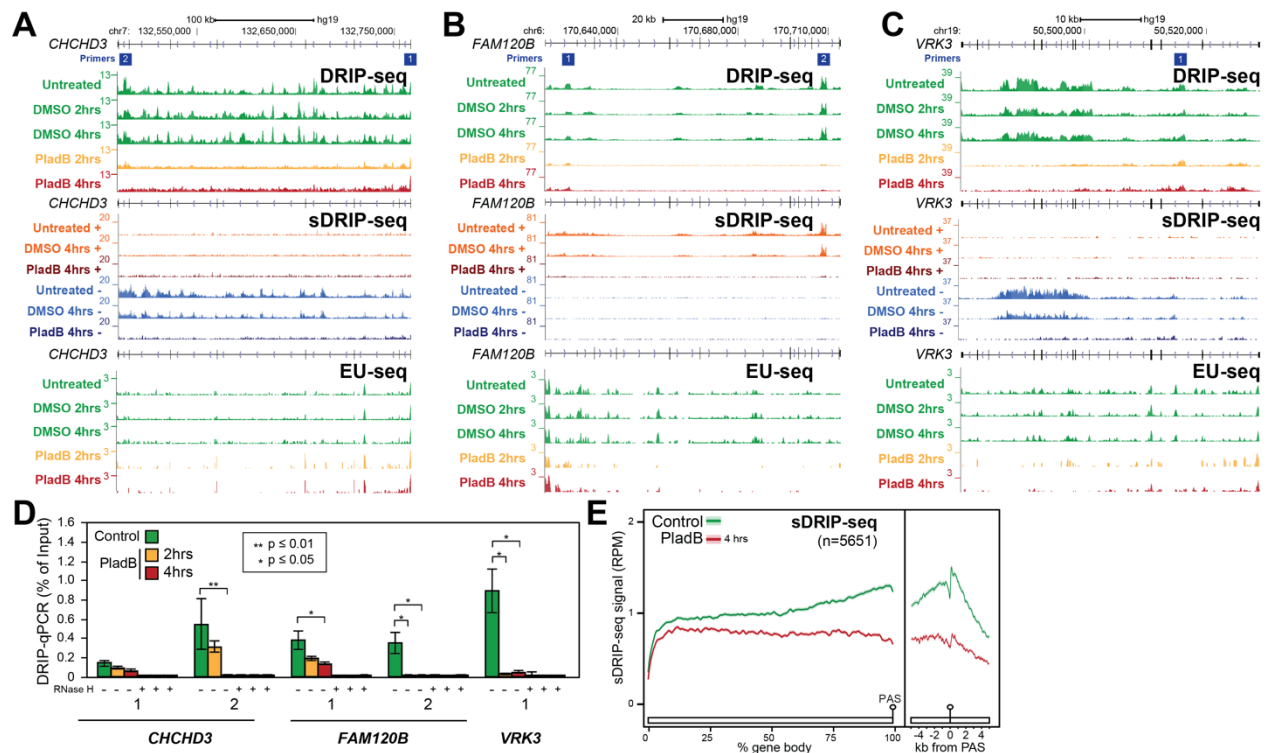
Figure 3.S4: A. Genome browser screenshot over the *CHCHD3* gene showing DRIP-seq signal (Top), sDRIP-signal (Middle), and EU-seq signal (Bottom) obtained from controls and PladB-treated K562 samples. R-loop loss is directional with transcription (colored bar). B. Genome browser screenshot over the *FAM120B* gene showing DRIP-seq signal (Top), sDRIP-signal (Middle), and EU-seq signal (Bottom) obtained from controls and PladB treated K562 samples. R-loop loss is directional with transcription

(colored bar). C. Genome browser screenshot over the *VRK3* gene showing DRIP-seq signal (Top), sDRIP-signal (Middle), and EU-seq (Bottom) obtained from controls and PladB treated K562 samples. D. Bar charts of DRIP-qPCR (as percent of input) for *CHCHD3, FAM120B, VRK3*. Each bar is the average of three independent experiments (shown with SE). Primer locations are indicated on each panel. E. Metaplot of sDRIP-seq signals over gene body (as % of gene length) and terminal regions (+/- 5kb from PAS) of RLL genes. Control and PladB-treated samples are color-coded as indicated. For each time point, the signal is shown as a trimmed mean (line) surrounded by the standard error (shaded).
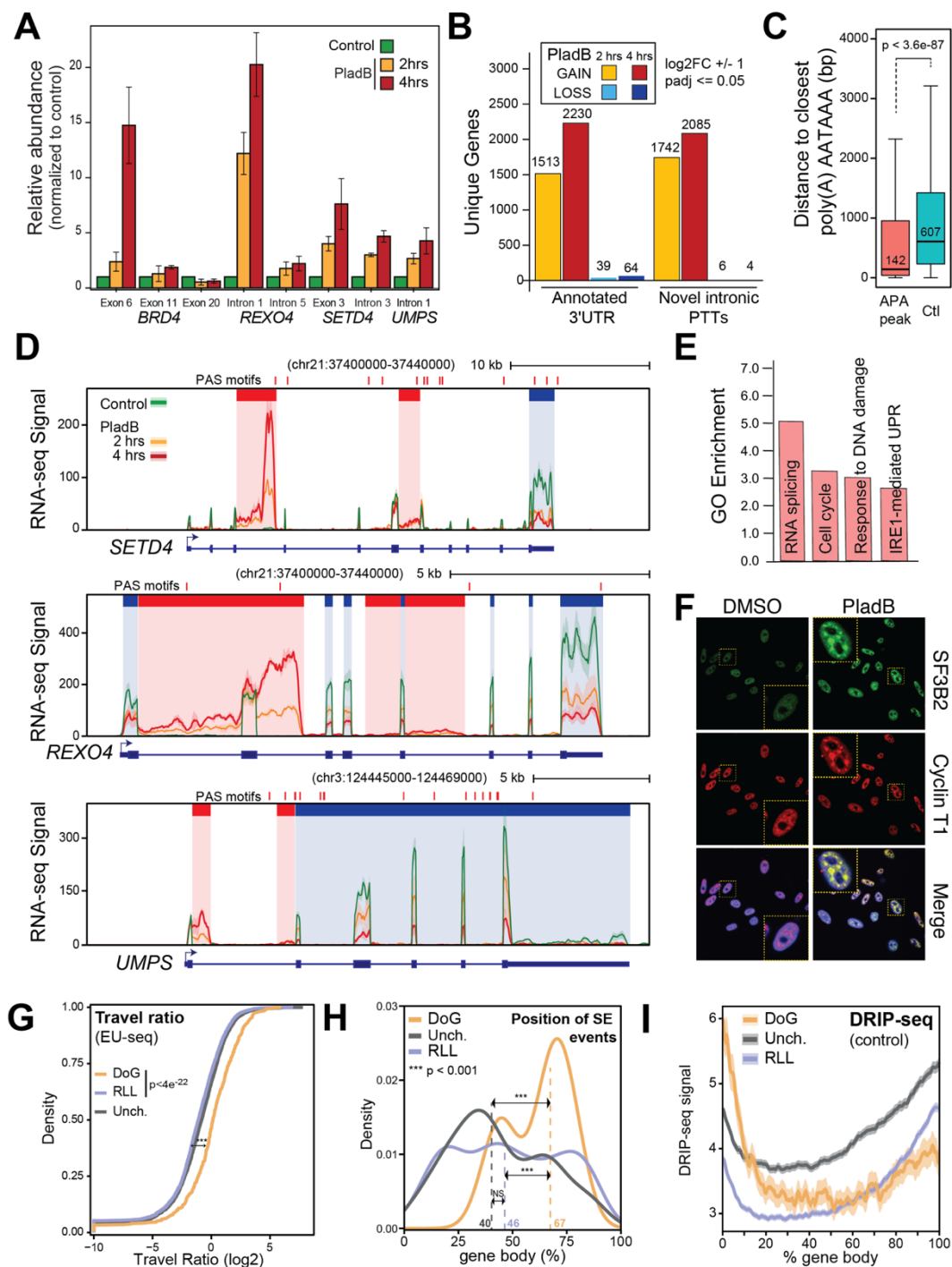
Figure 3.S5: A. Bar chart of RT-qPCR (as fold change over control) at indicated PladB-induced APA sites loci after PladB treatment. Each bar represents the average of three independent experiments (shown with SE). **B.** Systematic annotations of putative APA sites 4 hours after PladB treatment. Events were broken down between arising over previously annotated 3´UTRs and novel intronic 3´-UTRs. **C.** Distance between

the end of novel intronic APA regions and annotated polyA motifs. Control distances were calculated after shuffling the position of the APA regions within the same genes. P-value was calculated using a Wilcoxon test. **D.** RNA-seq signal for *REXO4*, *SETD4*, and *UMPS* under control and PladB-treated conditions. Highlighted in red and blue are sites of significantly increased or decreased APA usage. **E.** Enriched gene ontologies for Plad-induced APA genes. **F.** Representative immunofluorescence microscopy images from DMSO- and PladB-treated (4 hours) HeLa cells analyzed for SF3B2 and Cyclin T1 distribution and DAPI. Cyclin T1 becomes relocalized to splicing speckles. **G.** Travel ratios for DoG, RLL and unaffected genes are plotted using EU-seq signals under control conditions. Data is shown as the trimmed mean (line) along the standard error (shaded). **H.** Distribution of skipped exon events along DoG, unaffected, and RLL genes is plotted along gene bodies 4 hours post PladB treatment. The dashed lines indicate the position of the median position of such events. All p-values were determined by a Wilcoxon Mann-Whitney test (one, two, and three stars indicate p-value of ≤ 0.05, 0.01, and 0.001, respectively; NS: Not Significant). **I.** Metaplot of DRIP-seq signals along DoG, unaffected, and RLL genes were plotted along gene bodies under control conditions. Data is shown as the trimmed mean (line) along the standard error (shaded).

A
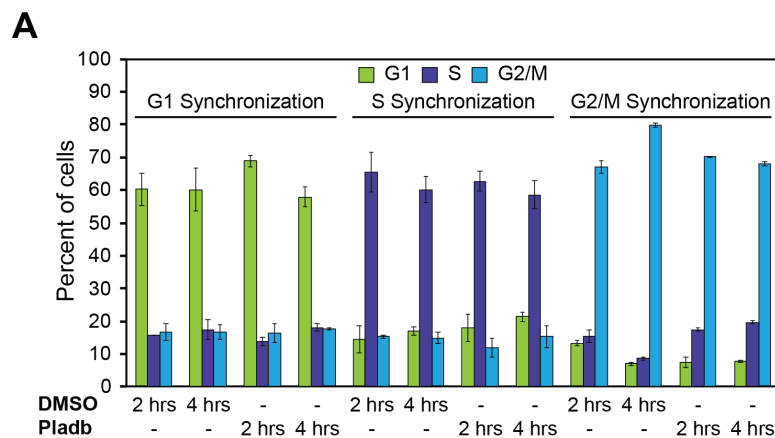
Figure 3.S6: A. Cell cycle distribution by PI staining: Bar plot representing the percentage of cells in G1, S, and G2/M in K562 cells. Bars represent an average of 4 independent measurements, Error bars represent SE.

3.8 Acknowledgements

3.9 Authors contributions

D.C.G performed all experiments with the assistance of L.A.S for R-loop mapping. S.R.H. performed computational data analysis with help from D.C.G. M.F. performed APA RT-qPCR experiments. F.C. conceived and supervised the project, analyzed data, and wrote the manuscript along with all authors.

CHAPTER IV: Concluding Remarks and Future Directions

This chapter includes text that was published in DNA Repair (Amst) Volume 106, 103183, October 2021.

The original manuscritpt has been modified to rename the figures based on the format of this dissertation. My contributions to the original manuscript included writing of the introduction and editing of the entire manscript. I also contributed to the figure making of all the figures in the original manuscript. Figure 4.1 was modified from the original manuscript.

4.1 Summary

Transcription is essential for gene expression, producing an RNA copy of the DNA template. Sometimes the RNA can become entangled with the DNA, creating R-loop structures. This project investigated whether splicing factors that control RNA processing can also regulate the formation of R-loops. Specifically, this work tested the model that inhibition of splicing causes excessive formation genome-wide over unsplice introns. In this work the global dynamic effects of PladB-mediated splicing inhibition on splicing, R-loop distribution, and transcription patterns were determined in order to clarify the temporal and positional relationships between splicing inhibition and Class II R-loop formation at high-resolution. Intron retention caused by SF3B1 inhibition does not trigger excessive R-loops. A subset of genes shows readthrough transcription and accompanying R-loop gains. R-loop gains and DNA damage were temporally and spatially uncoupled. Overall, SF3B1 inhibition causes broad reduction in nascent transcription and dominant R-loop loss. This project provided key insights into the inter-relationship between co-transcriptional splicing and Class II R-loop formation and their impact on transcriptional dynamics under stress conditions.

4.2 PladB-SF3B1 mediated splicing inhibition and its impact on Transcription

Recent studies have implicated the splicing process in transcriptional regulation and in transcriptional pausing (Gehring and Roignant, 2021; Herzel et al., 2017). For example, there is increased pausing at terminal exons for genes that are efficiently spliced in yeast (Alexander et al., 2010). When mutations are introduced into the branch point sequence (BPS) or U2 snRNA, there is an increase in RNA Pol II ChIP signal over introns (Chathoth et al., 2014), and RNA Pol II pause sites have been mapped to the 5′ SSs and 3′ SSs (Alexander et al., 2010; Harlen et al., 2016), which suggests RNA Pol II undergoes pausing in direct response to spliceosome assembly. What triggers this type of transcription elongation checkpoint to allow for assembly is unclear, but it could be that PladB is triggering the activation of this checkpoint.

From my work, we can conclude that the SF3b complex and U2 snRNP are critical for maintaining proper transcription elongation. Studies have shown that parts of the U2 snRNP are in close contact with the CTD of RNA Pol II during transcription (Hsin and Manley, 2012; Jacobs et al., 2004). As previously described, U2 snRNA base pairs with the BPS during spliceosome assembly. Within the B^{act} spliceosome complex SF3b proteins are in direct contact with the pre-mRNA upstream and downstream of the BPS and keeps the BPS away from the catalytic center. *In vitro* evidence suggests that initial spliceosome assembly is not completely impaired in the presence of PladB; U1 and U2 snRNP are able to bind to form the A complex, but unable to proceed with B complex formation (Effenberger et al., 2017; Finci et al., 2018; Jacobs et al., 2004). If this holds true *in vivo*, then U1 and U2 snRNPs might remain stuck or bound to the nascent mRNA transcript in the presence of PladB. These complexes are not able to be recycled and maintain close proximity with the nascent mRNA transcript. Since the conformational changes required for splicing to take place are blocked, this could result in the activation of a transcription elongation checkpoint to slow down RNA Pol II. Independent of how this checkpoint is activated, the outcome resulted in global reduction in nascent transcription. These transcriptional dynamics were driven by 1) premature transcription termination as evidence from the increase in intronic APA usage, and 2) global increase in promoter-proximal pausing. This cross talk between splicing and transcription could be rationalized in that it may ensure that the cell reduces efforts in transcribing nonfunctional/unspliced RNAs until splicing can be restored.

4.3 PladB and Other Stresses cause read-through Transcription

The phenomenon of transcription read-through, or DoG transcription, is not unique to PladB treatment. Several studies have shown that different stress conditions and viral infections result in readthrough transcription. Vilborg et al. (2015) were one of the first groups to show that osmotic stress trigged DoG transcription and generated long non-coding transcripts (>45 kb) past polyA signals (Vilborg et al., 2015). DoG transcripts also form under heat-shock and oxidative stress conditions (Vilborg et al., 2017; Vilborg and Steitz, 2017) and are nuclear-retained and chromatin-bound, as expected for R-loops.

Viral infections also impact transcriptional elongation and trigger DoG transcription and opening of the chromatin architecture (Bauer et al., 2018; Heinz et al., 2018). DoG transcription has also been characterized for senescence (Muniz et al., 2017) and in renal carcinoma cells (Grosso et al., 2015). In addition, these studies have determined that the genes that undergo DoG transcription vary depending on the stress condition and cell type. This suggests that DoG transcription is not random and that there could be an underlying biological function for this type of relatively novel phenotype.

The exact function of DoG transcription is unclear. Vilborg et al. (2017) proposed that DoG transcripts ensure that the nuclear architecture is maintained under stressed conditions (Vilborg et al., 2015; Vilborg et al., 2017; Vilborg and Steitz, 2017). In the context of viral infections, it was proposed that DoG transcription ensures that RNA Pol II remains bound to nuclear DNA and prevents it from becoming accessible to the viral machinery. More recent studies have proposed that stress-induced readthrough transcription may help regulate gene expression *via* post-transcriptional processing. For example, this may occur by altering the local chromatin landscape around their neighboring genes to facilitate responses throughout stress (Rosa-Mercado et al., 2021). Readthrough transcription can also result in maintenance of gene expression by maintaining the expression of the genes located directly downstream. Given the relative novelty of readthrough transcription, it is likely that more stresses cause defects in transcription termination. Further studies will decipher if there is an exact function to readthrough transcription or if it is the result of global disruptions in nuclear homeostatis under stress conditions.

At the root of DoG transcription is a failure in transcription termination, though what is impairing transcription termination under stress conditions is unclear. Transcription termination involves a multitude of different factors, such as CPSF and XRN2, and several studies have reported that depletion of these factors causes readthrough transcription (Eaton et al., 2018; Eaton et al., 2020; Fong et al., 2015). For example, depletion of CPSF subunit WDR33 resulted in transcription readthrough as a result of inefficient pre-mRNA cleavage and decreased transcription termination, suggesting that termination is coupled to proper mRNA cleavage (Teloni et al., 2019). It is important to point out that, unlike stress-induced

readthrough transcription, the effects of depleting this protein causes global increases in readthrough transcription compared to a subset of genes that are stress condition specific.

Recently, Rosa-Mercado et al. (2021) showed that osmotic stress causes the dissociation of the Integrator complex from RNA Pol II at genes that undergo readthrough transcription (n=590) and that depletion of the catalytic subunit of the Integrator complex, Int11, also resulted in readthrough transcription of hundreds of genes (n=348) (Rosa-Mercado et al., 2021). There was a partial overlap in the number of DoG genes between osmotic stress and Int11 depleted cells, suggesting that Integrator may have an indirect role in regulating DoG biogenesis. The observation that PladB treatment results in readthrough transcription and DoG R-loop formation offers the possibility that splicing disruptions, in part, may mechanistically underlie the formation of DoG transcripts. In support of this, viral infections that cause readthrough transcription negatively affect splicing and causes an increase in intron retention (Bauer et al., 2018). In addition, heat-shock and osmotic stress disrupt splicing. During heat shock splicing of mRNAs is inhibited (Shalgi et al., 2014), raising the possibility that the readthrough transcription that occurs under heat-shock can in part be driven by loss of SF3B1 function, similar to what is observed for PladB. While there is evidence that heat-shock and PladB may share a similar mechanism for readthrough transcription, it is more likely that there is more than a singular mechanism that drives readthrough transcription in different stress conditions. Additional work is required to understand how cells selectively induce DoGs from a subset of genes. For example, studies could focus on determining which accessory factors interact with RNA Pol II, as it transcribes through the DoG regions to maintain productive elongation. One could hypothesize that a new unidentified regulatory factor binds to RNA Pol II near the transcription termination site under stress conditions is critical for maintaining readthrough transcription.

A general model that can be proposed is that a stress condition can trigger termination failure through different modalities, such as by targeting XRN2, CPSF, Integrator, or SF3B1 (and other splicing factors), resulting in readthrough transcription. The observation that PladB DoG genes form R-loops suggests that the phenomenon of DoG transcription may be linked to R-loop formation under a variety of stresses. Further studies will be needed to determine if R-loops form under heat-shock or the other stress

conditions that cause readthrough and what the functional implications of DoG R-loop formation are. For example, DoG transcription upon viral infection has been further linked to large scale chromatin opening throughout the DoG region (Heinz et al., 2018). This raises the possibility that R-loops, which include a rigid A-form-like RNA:DNA hybrid, cause chromatin decondensation by preventing nucleosome wrapping or deposition. Methods that probe chromatin accessibility could be used to determine if all stress conditions result in opening of the chromatin. In addition, studies could explore the elongation rates of RNA Pol II through the DoG regions. Based on the increased R-loop formation of PladB DoG genes suggests that RNA Pol II may be transcribing at a slower rate (Sanz et al., 2016) through these intergenic regions.

What unifies all these stress responses, in addition to a subset of genes undergoing DoG transcription, is that they disrupt transcription elongation in a similar manner to PladB: there is a decrease in transcription elongation driven by premature transcription termination and an increase in RNA Pol II pausing. Maintaining active gene expression during a stress condition would potentially result in the production of nonfunctional/unspliced RNAs. It therefore makes sense that the cell may prioritize shutting down transcription of most genes within a short time frame (1 hour – 4 hours) to alleviate the burden of RNA quality surveillance pathways. In addition, premature transcription termination *via* APA could result in the production of RNA and protein isoforms that may be indirectly involved in driving the stress response (Berkovits and Mayr, 2015; Mayr, 2019; Williamson et al., 2017). Further studies will determine how protein homeostasis is altered by these transcriptional changes under each stress condition, will the goal of identifying the key factors involved in activating the "transcription elongation checkpoint" (described in 4.2) to slow down RNA Pol II.

4.4 Harmful R-loops and genome instability

Excessive R-loop formation was hypothesized early on to be responsible for the increased γH2AX deposition observed upon splicing factor depletion or mutation (Li and Manley, 2006; Li et al., 2005; Paulsen et al., 2009). We used DRIP-seq to profile Class II R-loops upon inhibition of the U2 spliceosome

SF3B1 subunit with PladB (Kotake et al., 2007; Yokoi et al., 2011). This treatment causes widespread intron retention as well as γH2AX accumulation and sensitivity to ATR inhibitors (Nguyen et al., 2018). Surprisingly, PladB treatment resulted in a dramatic genome-wide R-loop loss (Castillo-Guzman et al., 2020). These ~400 DoG genes represent the only detected increase in R-lop signal captured by DRIP/sDRIP, defining a class of *de novo* excessive R-loops (Castillo-Guzman et al., 2020). Importantly, the γH2AX accumulation caused by PladB was significantly delayed compared to *de novo* R-loop accumulation and was not spatially enriched over regions of increased R-loops (Castillo-Guzman et al., 2020). It therefore appears that excessive Class II R-loops generated during splicing inhibition did not associate with DNA damage events. More broadly, DNA damage events induced by U2 spliceosome inhibition occurred against the backdrop of a dramatic loss of Class II R-loops.

Harmful R-loops, whether they correspond to Class I or Class II R-loops, should in principle be revealed using appropriate genomic mapping techniques. It is therefore worth reviewing studies that credibly examined global R-loop patterns in cellular models of genome instability for any evidence that might clarify the nature of such structures and their spatiotemporal relationship to phenomena associated with DNA damage.

Class II R-loops and genome instability, an elusive connection.

DRIP-type approaches have been used in a variety of cellular models. DNA topoisomerase I (Top I) is widely thought to suppress R-loop formation by relaxing the R-loop-favorable negative superhelicity that propagates upstream of the active transcription machinery (Kouzine et al., 2004; Pommier et al., 2016). Long-term Top I depletion in human HeLa cells leads to elevated DNA damage and globally slower replication fork progression due to transcription-replication conflicts (Tuduri et al., 2009). Following up on this work, Promonet *et al.* (2020) mapped R-loop distributions, DNA breaks, and the location of DNA damage markers including phosphorylated RPA, a replicative stress marker, and γH2AX in Top I-depleted cells (Promonet et al., 2020). In control cells, R-loops were observed broadly over promoter distal regions,

gene bodies, and terminal regions, consistent with the distribution of elongation-associated R-loops. By contrast, phosphorylated RPA accumulation was only observed over the terminal regions of expressed genes that are replicated in a head-on (HO) orientation relative to transcription. This indicates that stalled forks marked by phosphorylated RPA occurred because of HO replication-transcription interactions, which coincide with naturally R-loop-rich terminal genic regions. The vast majority of Class II R-loops therefore do not interfere with DNA replication under normal conditions (Promonet et al., 2020). In Top I-depleted cells, R-loops showed a slight increase over terminal regions which was accompanied by increased γH2AX and DSB formation. Importantly, replication fork speeds were uniformly reduced by 30-40% in Top I-depleted cells, even though R-loop increases were minor and localized to the TTS. Thus, it is unlikely that "excessive" R-loop formation can account for the global replication slowdown. Instead, it was proposed that the stalled forks that naturally occur at a subset of HO genes are further challenged in the absence of Top I, leading to fork collapse, DSBs, and the activation of the ATR kinase to slow S phase progression globally (Promonet et al., 2020). Overall, this study suggests that Class II R-loops may not be directly involved in events of genome instability even upon Top I depletion.

Interestingly, RNase H1 over-expression was able to suppress the slow replication fork phenotype observed upon Top I depletion (Promonet et al., 2020; Tuduri et al., 2009). To account for this, Promonet et al. (2020) proposed that RNase H1 may degrade RNA:DNA hybrid structures that form over stalled forks and prevent fork rescue or remodeling. Through this activity, RNase H1 was suggested to reduce ATR activation and thereby counteract a global replication slowdown at undamaged forks, consistent with prior observations (Mutreja et al., 2018; Seiler et al., 2007). This is an exciting possibility that needs to be further tested. It also suggests that the sensitivity of a phenotype to RNase H1 over-expression may not necessarily implicate co-transcriptional R-loops and could instead reflect a novel effect of RNase H1 on replication fork rescue. Another possibility could be that RNase H1 is acting on other types of RNA:DNA hybrids not associated with Class II R-loops.

A number of studies provide qualified support for the notion that increased formation of Class II R-loops triggers or enhances genome instability phenomena. Stork et al. (2016) showed that addition of the hormone estrogen (E2) to breast cancer cells triggers rapid expression of E2-responsive genes and an increase in elongation-associated R-loops over these targets (Stork et al., 2016). E2 treatment also causes rapid cellular proliferation and increased deposition of the γH2AX DNA damage marker during S phase, indicating that R-loops may be driving this response. Indeed, rearrangements observed in breast tumors were enriched over E2-responsive genes. However, proximity ligation assays using S9.6 and γH2AX antibodies suggested that the majority of DNA damage events induced by E2 were located at a distance from R-loops (Stork et al., 2016). An alternative source of endogenous damage could come from the introduction of DSBs by Topoisomerase 2 beta at promoters, where it functions to facilitate transcription initiation in response to estrogens (Ju et al., 2006; Morimoto et al., 2019). Gorthi et al., (2018) showed that Ewing sarcoma cells driven by the EWS-FLI1 oncoprotein display globally elevated transcription levels and R-loop loads. These tumors were additionally characterized by increased replicative stress as well as sensitivity to ATR inhibitors and genotoxic agents (Gorthi et al., 2018). However, the direct involvement of R-loops in these responses remains unclear. Evidence suggests that the hyper-transcription caused by EWS-FLI1 results in the sequestration of BRCA1 with elongating RNA polymerase complexes, phenocopying a BRCA1 deficiency, and leading to a DNA damage response (DDR) deficiency (Gorthi et al., 2018). Two recent studies analyzed the impact of knocking out or depleting subunits of mammalian SWI/SNF chromatin remodeling complexes. Knockout of *PBRM1*, encoding the BAF180 subunit of the polybromo-associated BAF complex (PBAF), led to increased γH2AX foci formation, replication stress, and DNA breaks (Chabanon et al., 2021). This was correlated with an increased genic R-loop burden measured by DRIP-seq and the involvement of R-loops was further suggested by the ability of RNase H1 to rescue many of the above phenotypes. However, the spatial overlap of such excessive R-loops with DNA damage events was not assessed. In addition, the authors noted that PRBM1 deficiency led to significant reductions in the protein levels of multiple genome stability factors, such as members of the Fanconi

Anemia complex and of the BLM helicase (Chabanon et al., 2021). Thus, the genome instability observed in the absence of PBRM1 may reflect an intrinsically reduced DDR capacity in addition to elevated R-loop levels. Depletion of the BRG1 subunit common to all SWI/SNF complexes led to an array of RNase H1-sensitive genome instability phenotypes, including global decrease in replication fork velocity (Bayona-Feliu et al., 2021). DRIP-based profiling of Class II R-loops confirmed an increased burden of genic R-loops over 3,200 loci enriched for BRG1 binding sites. Analysis of transcription-replication conflicts showed that HO conflicts were marked by elevated DNA damage markers γH2AX and FANCD2, when co-directional (CD) conflicts were not, even though the R-loop levels observed at HO and CD regions were not significantly different (Bayona-Feliu et al., 2021). Thus, consistent with earlier statements, many R-loops outside of HO conflicts do not associate with DNA damage markers. HO interactions between transcription and replication, however, are particularly responsible for stalled forks and DNA damage. Whether the R-loops observed over these regions are causally involved in further enhancing the fragility of HO replication-transcription interactions currently rests on the interpretation of the sensitivity of instability phenotypes to RNase H1 over-expression. As discussed above, RNase H1 binding sites mapped by R-ChIP mostly map to promoter proximal regions while Class II R-loops mapped by DRIP approaches primarily map to transcribed gene bodies. The deployment of R-ChIP approaches in models of genome instability such as those described above may clarify whether RNase H1 can gain access to new R-loop subsets under these conditions or if it can be recruited directly to stalled forks, as suggested (Promonet et al., 2020).

Overall, the association of Class II R-loops and genome instability phenomena remains elusive. Part of the issue is that very few studies have both mapped R-loops and analyzed the presence of DNA breaks at sufficient resolution to reach definitive conclusions regarding the roles of R-loops as causes of genome instability. In addition, the observations that instability phenomena have been linked to global R-loop increases and losses in various models suggests that instability may arise through a variety of mechanisms. Indirect effects linked to reduced DDR responses under pathological conditions further complicate matters. To date, no specific R-loop subset has been associated directly with events of DNA

breakage at high resolution. Furthermore, the distinguishing molecular features of harmful R-loops remain to be defined, an important task given the consensus finding that most Class II R-loops do not cause instability even under altered conditions.

Class I R-loops as harmful R-loops candidates.

Overexpression of nuclear RNase H1 suppresses a diversity of genome instability phenotypes. Common sense dictates that for RNase H1 to mediate these effects, it needs to gain access to the loci causing these altered phenotypes in the first place. Based on the few available datasets where RNase H1 genomic binding sites were mapped in human cell lines, these loci predominantly correspond to Class I R-loops that form at the promoter-proximal regions of paused promoters (Chen et al., 2018; Chen et al., 2017). By contrast, most Class II R-loops are not bound by RNase H1 *in vivo*. It is therefore difficult to conceive how RNase H1 might relieve instability phenotypes if these were driven by elongation-associated R-loops.

One interpretation of the data is therefore that harmful R-loops predominantly correspond to Class I R-loops. This model is attractive for several reasons. It has been well documented that the deleterious effects of harmful R-loops are linked to conflicts with the replication machinery (Garcia-Muse and Aguilera, 2016; Gomez-Gonzalez and Aguilera, 2019; Hamperl and Cimprich, 2016; Zeman and Cimprich, 2014). Class I R-loops possess multiple properties that might make them much more formidable replication obstacles than Class II R-loops. First, they are proposed to arise at high frequencies compared to Class II R-loops, consistent with the much greater RNA Pol II density at paused promoters compared to gene bodies. Second, Class I R-loops consistently occur in a narrow near-TSS genomic window, while Class II R-loops can occur almost anywhere throughout gene bodies. Thus, encounters between replication forks and R-loops will be more likely to involve Class I R-loops, and these encounters are expected to be focused over paused promoters. Third, one major difference between Class I and Class II R-loops is that the former is associated with a paused, and possibly backtracked, RNA Pol II transcription machinery (Sheridan et al., 2019). By extension, Class I R-loops will also be proximal to the large general transcription factor (GTF) complexes that recruit RNA Pol II to promoter sequences (Verger et al., 2021), limiting the range that Class

I R-loops can extend and increase the frequency at which they form. By contrast, Class II R-loops occur during elongation behind an actively translocating RNA Pol II, far away from GTF complexes. One can even envision that RNA Pol II has moved away from an R-loop once R-loop extension is terminated.

Recent *in vitro* work suggests that R-loops by themselves do not represent a strong impediment for the *E. coli* replication machinery, while the presence of transcription complexes led to potent blockages, particularly in the head-on orientation (Bruning and Marians, 2020). In addition, R-loop-anchored transcription complexes arrested at UV lesions were proposed to represent the main cause of head-on replication blocks in RNase H-deficient *E. coli* mutants (Kouzminova and Kuzminov, 2021). Class I R-loops, associated with paused RNA Pol II complexes, therefore represent attractive "harmful" R-loop candidates. We note that human genes are generally thought to be replicated co-directionally from origins located upstream of promoter regions (Petryk et al., 2016). Thus, one would expect encounters between Class I R-loops and replication forks to be mostly co-directional. These interactions, while potentially harmful (Hamperl et al., 2017), are thought to play a lesser role compared to head-on conflicts in driving genome instability (Gomez-Gonzalez and Aguilera, 2019; Hamperl and Cimprich, 2016). It remains possible, however, that Class I R-loops might also arise from antisense transcripts that frequently originate from promoters, setting up head-on clashes with incoming forks. In addition, recent evidence suggests that antisense-associated promoter regions display delayed replication characteristics that may factor in possible fragility events (Wang et al., 2021a). Studies aimed at dissecting transcription-replication encounters and their intersection with Class I R-loop formation, will be important to further delineate the role of R-loops in genome fragility.

A growing number of studies provide support to the notion that Class I R-loops increase in frequency under pathological instances. Using R-ChIP, Chen and colleagues profiled Class I R-loops in HEK293T cells harboring mutations in the splicing factors SRSF2, U2AF1, and U2AF2 (Chen et al., 2018). Mutations in SRSF2 and U2AF1 caused cellular growth defects, γH2AX induction, and replicative stress as evidenced by slower replication forks and ATR activation. Such defects could be at least partially

alleviated by RNase H1over-expression. R-ChIP revealed increased Class I R-loops over promoter regions. SRSF2 and U2AF1 mutations also caused increased RNA Pol II promoter-pausing (Chen et al., 2018), further linking promoter-proximal pausing and Class I R-loop formation. Pharmacological splicing inhibition also triggers increased promoter pausing (Caizzi et al., 2021; Castillo-Guzman et al., 2020; Sousa-Luis et al., 2021) suggesting that Class I R-loops may generally increase under these conditions. Release from the promoter-proximal pause is regulated by the pTEFb complex which is recruited in part via the general BRD4 co-activator protein (Kwak and Lis, 2013). Loss of BRD4 function triggers global transcriptional pausing (Muhar et al., 2018). Pharmacological inhibition of BRD4 or depletion of BRD4 cause S phase-dependent γH2AX deposition, DSB formation, and slow replication fork progression in a manner that was compensated by RNase H1 expression (Edwards et al., 2020; Lam et al., 2020). Performing R-ChIP, Edwards *et al.*, (2020) showed that BRD4 loss of function caused global increase in Class I R-loops over promoters, as expected, and also in gene bodies, which is counter-intuitive given the reduction of the elongating form of RNA Pol II (Edwards et al., 2020). While none of studies cited above mapped DSBs and therefore did not directly address the connection between R-loops and DNA breakage, they support the notion that Class I R-loops may represent harmful obstacles to replication progression and a source of DNA damage.

More broadly, we propose that defects in mRNA processing, particularly splicing dysfunction, may feedback on transcription by increasing promoter pausing and therefore, increasing Class I R-loops. The increased burden of paused transcriptional complexes anchored to Class I R-loops is expected to enhance replication-transcription conflicts over promoter regions and may lead to replication stress and R-loop-induced DNA damage. Given that increased promoter pausing associates with reduced transcription initiation (Gressel et al., 2019; Gressel et al., 2017; Shao and Zeitlinger, 2017), the model further predicts that Class II R-loops may undergo progressive losses as a result of reduced elongation, as observed upon U2 spliceosome inhibition (Castillo-Guzman et al., 2020).

4.5 Concluding Remarks

Overall, this work highlights the complexity of cellular stress response pathways. If we consider inhibition of splicing a stress condition, then it makes sense why the cell undergoes global changes at the transcriptional level. Splicing inhibition *via* PladB treatment triggers global decrease in nascent transcription and a global increase in RNA Pol II promoter pausing. The first ensures that the cell does not waste valuable resources to produce transcripts that are not viable, and the second response ensures that the cell is primed and ready to jump start the transcription downregulated genes when the stressor is relieved. It is clear that in addition to U1 snRNP, U2 snRNP also has a clear role in regulating transcription elongation. What remains unclear is what is the mechanism causing DNA damage under these conditions. One possibility is that perturbations in splicing cause DDR genes to be downregulated. Downregulation of DDR genes could sensitize cells to basal levels of DNA damage the cell experiences through normal cellular activity and metabolism (Figure 4.1). Both PRMT1 inhibition (Giuliani et al., 2021) and E7107 splicing inhibitor (Han et al., 2022) were proposed to drive increased DNA damage in a similar manner. Another possibility as described above is that the paused RNA Pol II itself or the formation of Class I R-loops could be contributing to genome instability (Figure 4.1).
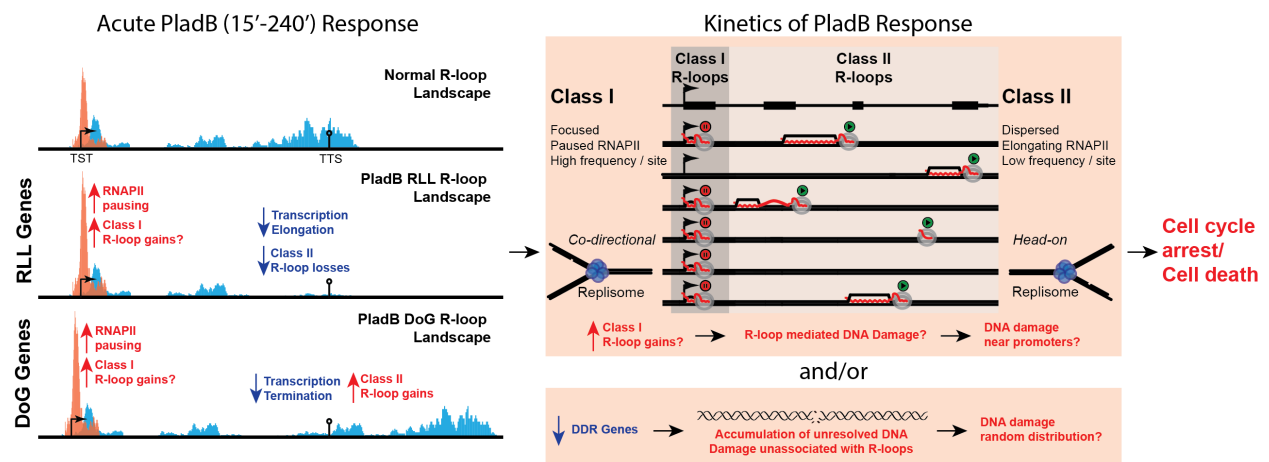
Figure 4.1 Kinetics of PladB response. PladB-SF3B1 mediated splicing inhibition caused **RNA Pol II pausing** in RLL and DoG genes. This increased paused is proposed to cause a focal increase in Class I R-loops. Class II R-loops by contrast were reduced as a result of **lower transcription initiation** and **increased termination**. Class I R-loops are suggested to frequently arise at promoters as a result of **promoter-proximal pausing**; PladB treatment is proposed to further increase their formation frequency. By contrast, elongation-associated Class II R-loops are spread throughout the gene body and therefore occur at much lower frequencies at any given site; PladB treatment will further reduce their frequency. Under such conditions, paused RNA Pol II complexes anchored to Class I R-loops may represent frequent obstacles to replication fork progression, causing transcription-replication conflicts and DNA damage focused over promoter regions. RNA is only depicted when in a R-loop bound state or exiting RNAP (full transcript not shown). Second model DNA damage can arise: global reduction in transcription may cause downregulation of DNA damage response (DDR) genes resulting in accumulation of DNA damage unassociated with R-loops.

The notion that RNA Pol II-driven R-loops can be broken into pausing-associated and elongation-associated structures hopefully serves to reconcile and rationalize the seemingly discordant results obtained through *ex vivo* and native R-loop mapping approaches. With the right methodologies now at hand, new investigations will test the proposal that Class I R-loops significantly contribute to genome instability phenotypes associated with RNA processing defects. We suggest that future work should include integrative strategies that reveal the distributions of both Class I and Class II R-loops under relevant cellular models of R-loop dysfunction in combination with DNA double-strand break profiling and nascent transcription analysis. Such integrative studies are the most likely to reveal the mechanisms that lead to genome destabilization when gene expression programs are deregulated. Given the importance of RNase H1 over-expression as a tool and its broad ability to suppress genomic stresses, it will also be essential to ascertain how suppression is mechanistically achieved. If Class I R-loops truly emerge as source of DNA damage under pathological conditions, it will become important to understand how RNase H1 activity at

paused promoters can lower transcription-replication conflicts. One possibility is that RNase H1, when over-expressed, facilitates the release of paused RNA Pol II into elongation (Sridhara et al., 2017). Alternatively, RNase H1 may mediate its effect by facilitating the processing and restart of stalled replication forks (Promonet et al., 2020) in the vicinity of paused promoters. Interestingly, recent evidence suggests that the persistence of RNA Pol II at TSSs characterized by strong antisense transcription prevents the timely replication of these loci until G2/M (Wang et al., 2021a). In agreement, mapping of transcription-replication interactions in murine B cells revealed that such interactions primarily occurred over regions characterized by bidirectional promoters and focal accumulation of markers of replicative stress and DNA breakage (St Germain et al., 2021). Future work will be required to tease out these interesting possibilities.

CHAPTER V: REFERENCES

Abbas, T., Keaton, M.A., and Dutta, A. (2013). Genomic instability in cancer. Cold Spring Harb Perspect Biol *5*, a012914.

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. Nature *438*, 460-465.

Adelman, K., La Porta, A., Santangelo, T.J., Lis, J.T., Roberts, J.W., and Wang, M.D. (2002). Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. Proc Natl Acad Sci U S A *99*, 13538-13543.

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature reviews Genetics *13*, 720-731.

Adkins, M.W., and Tyler, J.K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. Molecular cell *21*, 405-416.

Aguilera, A. (2005a). Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. Current opinion in cell biology *17*, 242-250.

Aguilera, A. (2005b). mRNA processing and genomic instability. Nat Struct Mol Biol *12*, 737-738.

Aguilera, A., and Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Molecular cell *46*, 115-124.

Alecki, C., Chiwara, V., Sanz, L.A., Grau, D., Arias Perez, O., Boulier, E.L., Armache, K.J., Chedin, F., and Francis, N.J. (2020). RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2. Nature communications *11*, 1781.

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

Alpert, T., Herzel, L., and Neugebauer, K.M. (2017). Perfect timing: splicing and transcription rates in living cells. Wiley Interdiscip Rev RNA *8*.

Arnold, M., Bressin, A., Jasnovidova, O., Meierhofer, D., and Mayer, A. (2021). A BRD4-mediated elongation control point primes transcribing RNA polymerase II for 3'-processing and termination. Molecular cell *81*, 3589-3603 e3513.

Aronica, L., Kasparek, T., Ruchman, D., Marquez, Y., Cipak, L., Cipakova, I., Anrather, D., Mikolaskova, B., Radtke, M., Sarkar, S.*, et al.* (2016). The spliceosome-associated protein Nrl1 suppresses homologous recombination-dependent R-loop formation in fission yeast. Nucleic acids research *44*, 1703-1717.

Aslanzadeh, V., Huang, Y., Sanguinetti, G., and Beggs, J.D. (2018). Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. Genome Res *28*, 203-213.

Azkanaz, M., Rodriguez Lopez, A., de Boer, B., Huiting, W., Angrand, P.O., Vellenga, E., Kampinga, H.H., Bergink, S., Martens, J.H., Schuringa, J.J.*, et al.* (2019). Protein quality control in the nucleolus safeguards recovery of epigenetic regulators after heat shock. Elife *8*.

Baejen, C., Torkler, P., Gressel, S., Essig, K., Soding, J., and Cramer, P. (2014). Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. Molecular cell *55*, 745-757.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. Cell *120*, 183-193.

Barroso, S., Herrera-Moyano, E., Munoz, S., Garcia-Rubio, M., Gomez-Gonzalez, B., and Aguilera, A. (2019). The DNA damage response acts as a safeguard against harmful DNA-RNA hybrids of different origins. EMBO reports *20*, e47250.

Bauer, D.L.V., Tellier, M., Martinez-Alonso, M., Nojima, T., Proudfoot, N.J., Murphy, S., and Fodor, E. (2018). Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell Rep *23*, 2119-2129 e2113.

Bayona-Feliu, A., Barroso, S., Munoz, S., and Aguilera, A. (2021). The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated transcription-replication conflicts. Nature genetics.

Belanger, K.G., and Kreuzer, K.N. (1998). Bacteriophage T4 initiates bidirectional DNA replication through a two-step process. Molecular cell *2*, 693-701.

Belotserkovskii, B.P., Soo Shin, J.H., and Hanawalt, P.C. (2017). Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. Nucleic acids research *45*, 6589-6599.

Belotserkovskii, B.P., Tornaletti, S., D'Souza, A.D., and Hanawalt, P.C. (2018). R-loop generation during transcription: Formation, processing and cellular outcomes. DNA repair.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nature reviews Genetics *15*, 163-175.

Bentley, D.L., and Groudine, M. (1986). A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells. Nature *321*, 702-706.

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L.*, et al.* (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell *150*, 53-64.

Berkovits, B.D., and Mayr, C. (2015). Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. Nature *522*, 363-367.

Bertram, K., Agafonov, D.E., Liu, W.T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Luhrmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. Nature *542*, 318-323.

Bhatia, V., Barroso, S.I., Garcia-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.

Boija, A., Mahat, D.B., Zare, A., Holmqvist, P.H., Philip, P., Meyers, D.J., Cole, P.A., Lis, J.T., Stenberg, P., and Mannervik, M. (2017). CBP Regulates Recruitment and Release of Promoter-Proximal RNA Polymerase II. Molecular cell *68*, 491-503 e495.

Bondarenko, V.A., Steele, L.M., Ujvari, A., Gaykalova, D.A., Kulaeva, O.I., Polikanov, Y.S., Luse, D.S., and Studitsky, V.M. (2006). Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. Molecular cell *24*, 469-479.

Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Geli, V., de Almeida, S.F., and Palancade, B. (2017). Introns Protect Eukaryotic Genomes from Transcription-Associated Genetic Instability. Molecular cell *67*, 608-621 e606.

Boswell, S.A., Snavely, A., Landry, H.M., Churchman, L.S., Gray, J.M., and Springer, M. (2017). Total RNA-seq to identify pharmacological effects on specific stages of mRNA synthesis. Nat Chem Biol *13*, 501-507.

Braberg, H., Jin, H., Moehle, E.A., Chan, Y.A., Wang, S., Shales, M., Benschop, J.J., Morris, J.H., Qiu, C., Hu, F.*, et al.* (2013). From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell *154*, 775-788.

Brugiolo, M., Herzel, L., and Neugebauer, K.M. (2013). Counting on co-transcriptional splicing. F1000Prime Rep *5*, 9.

Bruning, J.G., and Marians, K.J. (2020). Replisome bypass of transcription complexes and R-loops. Nucleic acids research *48*, 10353-10367.

Cadoret, J.C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc Natl Acad Sci U S A *105*, 15837-15842.

Cai, S., Bai, Y., Wang, H., Zhao, Z., Ding, X., Zhang, H., Zhang, X., Liu, Y., Jia, Y., Li, Y.*, et al.* (2020). Knockdown of THOC1 reduces the proliferation of hepatocellular carcinoma and increases the sensitivity to cisplatin. J Exp Clin Cancer Res *39*, 135.

Caizzi, L., Monteiro-Martins, S., Schwalb, B., Lysakovskaia, K., Schmitzova, J., Sawicka, A., Chen, Y., Lidschreiber, M., and Cramer, P. (2021). Efficient RNA polymerase II pause release requires U2 snRNP function. Molecular cell *81*, 1920-1934 e1929.

Cardiello, J.F., Goodrich, J.A., and Kugel, J.F. (2018). Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination. Molecular and cellular biology *38*.

Carles-Kinch, K., and Kreuzer, K.N. (1997). RNA-DNA hybrid formation at a bacteriophage T4 replication origin. J Mol Biol *266*, 915-926.

Carrasco-Salas, Y., Malapert, A., Sulthana, S., Molcrette, B., Chazot-Franguiadakis, L., Bernard, P., Chedin, F., Faivre-Moskalenko, C., and Vanoosthuyse, V. (2019). The extruded non-template strand determines the architecture of R-loops. Nucleic acids research.

Carvalho, T., Martins, S., Rino, J., Marinho, S., and Carmo-Fonseca, M. (2017). Pharmacological inhibition of the spliceosome subunit SF3b triggers exon junction complex-independent nonsense-mediated decay. J Cell Sci *130*, 1519-1531.

Castillo-Guzman, D., and Chedin, F. (2021). Defining R-loop classes and their contributions to genome instability. DNA repair *106*, 103182.

Castillo-Guzman, D., Hartono, S.R., Sanz, L.A., and Chédin, F. (2020). SF3B1-targeted Splicing Inhibition Triggers Global Alterations in Transcriptional Dynamics and R-Loop Metabolism. bioRxiv, 2020.2006.2008.130583.

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

Chabanon, R.M., Morel, D., Eychenne, T., Colmet-Daage, L., Bajrami, I., Dorvault, N., Garrido, M., Meisenberg, C., Lamb, A., Ngo, C.*, et al.* (2021). PBRM1 Deficiency Confers Synthetic Lethality to DNA Repair Inhibitors in Cancer. Cancer research.

Chakraborty, P., Huang, J.T.J., and Hiom, K. (2018). DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. Nature communications *9*, 4346.

Chan, Y.A., Hieter, P., and Stirling, P.C. (2014). Mechanisms of genome instability induced by RNA-processing defects. Trends in genetics : TIG *30*, 245-253.

Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. Molecular cell *53*, 779-790.

Chedin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. Trends in genetics : TIG *32*, 828-838.

Chedin, F., and Benham, C.J. (2020). Emerging roles for R-loop structures in the management of topological stress. J Biol Chem *295*, 4684-4695.

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

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

Chen, H.T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. Nat Struct Mol Biol *14*, 696-703.

Chen, J.Y., Zhang, X., Fu, X.D., and Chen, L. (2019). R-ChIP for genome-wide mapping of R-loops by using catalytically inactive RNASEH1. Nat Protoc *14*, 1661-1685.

Chen, L., Chen, J.Y., Huang, Y.J., Gu, Y., Qiu, J., Qian, H., Shao, C., Zhang, X., Hu, J., Li, H.*, et al.* (2018). The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. Molecular cell *69*, 412-425 e416.

Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H.*, et al.* (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. Molecular cell *68*, 745-757 e745.

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 *22*, 999-1007.

Chen, Z.A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J.C., Nilges, M*., et al.* (2010). Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. EMBO J *29*, 717-726.

Cheng, B., Li, T., Rahl, P.B., Adamson, T.E., Loudas, N.B., Guo, J., Varzavand, K., Cooper, J.J., Hu, X., Gnatt, A*., et al.* (2012). Functional association of Gdown1 with RNA polymerase II poised on human genes. Molecular cell *45*, 38-50.

Cheung, A.C., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature *471*, 249-253.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: the exosome. Biochimica et biophysica acta *1829*, 552-560.

Cohen, S., Puget, N., Lin, Y.L., Clouaire, T., Aguirrebengoa, M., Rocher, V., Pasero, P., Canitrot, Y., and Legube, G. (2018). Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. Nature communications *9*, 533.

Compe, E., and Egly, J.M. (2016). Nucleotide Excision Repair and Transcriptional Regulation: TFIIH and Beyond. Annual review of biochemistry *85*, 265-290.

Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev *33*, 960-982.

Cortazar, M.A., Sheridan, R.M., Erickson, B., Fong, N., Glover-Cutter, K., Brannan, K., and Bentley, D.L. (2019). Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. Molecular cell *76*, 896-908 e894.

Costantino, L., and Koshland, D. (2015). The Yin and Yang of R-loop biology. Current opinion in cell biology *34*, 39-45.

Cozzarelli, N.R., Cost, G.J., Nollmann, M., Viard, T., and Stray, J.E. (2006). Giant proteins that move DNA: bullies of the genomic playground. Nature reviews Molecular cell biology *7*, 580-588.

Cramer, P. (2004). Structure and function of RNA polymerase II. Adv Protein Chem *67*, 1-42.

Cretu, C., Agrawal, A.A., Cook, A., Will, C.L., Fekkes, P., Smith, P.G., Luhrmann, R., Larsen, N., Buonamici, S., and Pena, V. (2018). Structural Basis of Splicing Modulation by Antitumor Macrolide Compounds. Molecular cell *70*, 265-273 e268.

Crossley, M.P., Bocek, M., and Cimprich, K.A. (2019a). R-Loops as Cellular Regulators and Genomic Threats. Molecular cell *73*, 398-411.

Crossley, M.P., Bocek, M., Hamperl, S., Swigut, T., and Cimprich, K.A. (2019b). qDRIP: Quantitative differential RNA:DNA hybrid immunoprecipitation sequencing. bioRxiv.

Crossley, M.P., Bocek, M.J., Hamperl, S., Swigut, T., and Cimprich, K.A. (2020). qDRIP: a method to quantitatively assess RNA-DNA hybrid formation genome-wide. Nucleic acids research *48*, e84.

Cugusi, S., Mitter, R., Kelly, G.P., Walker, J., Han, Z., Pisano, P., Wierer, M., Stewart, A., and Svejstrup, J.Q. (2022). Heat shock induces premature transcript termination and reconfigures the human transcriptome. Molecular cell.

D'Alessandro, G., Whelan, D.R., Howard, S.M., Vitelli, V., Renaudin, X., Adamowicz, M., Iannelli, F., Jones-Weinert, C.W., Lee, M., Matti, V*., et al.* (2018). BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. Nature communications *9*, 5376.

Dahlberg, O., Shilkova, O., Tang, M., Holmqvist, P.H., and Mannervik, M. (2015). P-TEFb, the super elongation complex and mediator regulate a subset of non-paused genes during early Drosophila embryo development. PLoS genetics *11*, e1004971.

Dangkulwanich, M., Ishibashi, T., Liu, S., Kireeva, M.L., Lubkowska, L., Kashlev, M., and Bustamante, C.J. (2013). Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism. Elife *2*, e00971.

Daniels, G.A., and Lieber, M.R. (1995). RNA:DNA complex formation upon transcription of immunoglobulin switch regions: implications for the mechanism and regulation of class switch recombination. Nucleic acids research *23*, 5006-5011.

Davidson, L., and West, S. (2013). Splicing-coupled 3' end formation requires a terminal splice acceptor site, but not intron excision. Nucleic acids research *41*, 7101-7114.

de Almeida, S.F., Garcia-Sacristan, A., Custodio, N., and Carmo-Fonseca, M. (2010). A link between nuclear RNA surveillance, the human exosome and RNA polymerase II transcriptional termination. Nucleic acids research *38*, 8015-8026.

Derti, A., Garrett-Engele, P., Macisaac, K.D., Stevens, R.C., Sriram, S., Chen, R., Rohl, C.A., Johnson, J.M., and Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. Genome Res *22*, 1173-1183.

Dey, A., Nishiyama, A., Karpova, T., McNally, J., and Ozato, K. (2009). Brd4 marks select genes on mitotic chromatin and directs postmitotic transcription. Molecular biology of the cell *20*, 4899-4909.

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

Donner, A.J., Ebmeier, C.C., Taatjes, D.J., and Espinosa, J.M. (2010). CDK8 is a positive regulator of transcriptional elongation within the serum response network. Nat Struct Mol Biol *17*, 194-201.

Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Molecular cell *77*, 985-998 e988.

Drolet, M., Broccoli, S., Rallu, F., Hraiky, C., Fortin, C., Masse, E., and Baaklini, I. (2003). The problem of hypernegative supercoiling and R-loop formation in transcription. Front Biosci *8*, d210-221.

Drolet, M., and Brochu, J. (2019). R-loop-dependent replication and genomic instability in bacteria. DNA repair *84*, 102693.

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

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. Genes Dev *18*, 1618-1629.

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev *32*, 127-139.

Eaton, J.D., Francis, L., Davidson, L., and West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Genes Dev *34*, 132-145.

Eaton, J.D., and West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! Trends in genetics : TIG *36*, 664-675.

Edwards, D.S., Maganti, R., Tanksley, J.P., Luo, J., Park, J.J.H., Balkanska-Sinclair, E., Ling, J., and Floyd, S.R. (2020). BRD4 Prevents R-Loop Formation and Transcription-Replication Conflicts by Ensuring Efficient Transcription Elongation. Cell Rep *32*, 108166.

Effenberger, K.A., Urabe, V.K., and Jurica, M.S. (2017). Modulating splicing with small molecular inhibitors of the spliceosome. Wiley Interdiscip Rev RNA *8*.

Eick, D., and Bornkamm, G.W. (1986). Transcriptional arrest within the first exon is a fast control mechanism in c-myc gene expression. Nucleic acids research *14*, 8331-8346.

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 genetics *10*, e1004716.

Epshtein, V., Cardinale, C.J., Ruckenstein, A.E., Borukhov, S., and Nudler, E. (2007). An allosteric path to transcription termination. Molecular cell *28*, 991-1001.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Molecular cell *36*, 593-608.

Feaver, W.J., Svejstrup, J.Q., Henry, N.L., and Kornberg, R.D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, 1103-1109.

Fica, S.M., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.C., Newman, A.J., and Nagai, K. (2017). Structure of a spliceosome remodelled for exon ligation. Nature *542*, 377-380.

Finci, L.I., Zhang, X., Huang, X., Zhou, Q., Tsai, J., Teng, T., Agrawal, A., Chan, B., Irwin, S., Karr, C.*, et al.* (2018). The cryo-EM structure of the SF3b spliceosome complex bound to a splicing modulator reveals a pre-mRNA substrate competitive mechanism of action. Genes Dev *32*, 309-320.

Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., and Bentley, D.L. (2015). Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. Molecular cell *60*, 256-267.

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev *28*, 2663-2676.

Forth, S., Sheinin, M.Y., Inman, J., and Wang, M.D. (2013). Torque measurement at the single-molecule level. Annu Rev Biophys *42*, 583-604.

Fourmann, J.B., Schmitzova, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.L., Fabrizio, P., and Luhrmann, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. Genes Dev *27*, 413-428.

Fuchs, G., Hollander, D., Voichek, Y., Ast, G., and Oren, M. (2014). Cotranscriptional histone H2B monoubiquitylation is tightly coupled with RNA polymerase II elongation rate. Genome Res *24*, 1572-1583.

Galbraith, M.D., Allen, M.A., Bensard, C.L., Wang, X., Schwinn, M.K., Qin, B., Long, H.W., Daniels, D.L., Hahn, W.C., Dowell, R.D.*, et al.* (2013). HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. Cell *153*, 1327-1339.

Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. Nature *537*, 197-201.

Garcia-Muse, T., and Aguilera, A. (2016). Transcription-replication conflicts: how they occur and how they are resolved. Nature reviews Molecular cell biology *17*, 553-563.

Garcia-Muse, T., and Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. Cell *179*, 604-618.

Gehring, N.H., and Roignant, J.Y. (2021). Anything but Ordinary - Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. Trends in genetics : TIG *37*, 355-372.

Geisberg, J.V., Moqtaderi, Z., and Struhl, K. (2020). The transcriptional elongation rate regulates alternative polyadenylation in yeast. Elife *9*.

Giannakakis, A., Zhang, J., Jenjaroenpun, P., Nama, S., Zainolabidin, N., Aau, M.Y., Yarmishyn, A.A., Vaz, C., Ivshina, A.V., Grinchuk, O.V.*, et al.* (2015). Contrasting expression patterns of coding and noncoding parts of the human genome upon oxidative stress. Sci Rep *5*, 9737.

Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I., and Chedin, F. (2013). GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. Genome Res *23*, 1590-1600.

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

Girard, C., Will, C.L., Peng, J., Makarov, E.M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K., and Luhrmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. Nature communications *3*, 994.

Giuliani, V., Miller, M.A., Liu, C.Y., Hartono, S.R., Class, C.A., Bristow, C.A., Suzuki, E., Sanz, L.A., Gao, G., Gay, J.P.*, et al.* (2021). PRMT1-dependent regulation of RNA metabolism and DNA damage response sustains pancreatic ductal adenocarcinoma. Nature communications *12*, 4626.

Gomez-Gonzalez, B., and Aguilera, A. (2019). Transcription-mediated replication hindrance: a major driver of genome instability. Genes Dev *33*, 1008-1026.

Gomez-Gonzalez, B., Garcia-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marin, A., Foiani, M., and Aguilera, A. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. EMBO J *30*, 3106-3119.

Gorthi, A., Romero, J.C., Loranc, E., Cao, L., Lawrence, L.A., Goodale, E., Iniguez, A.B., Bernard, X., Masamsetti, V.P., Roston, S.*, et al.* (2018). EWS-FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. Nature *555*, 387-391.

Gottipati, P., and Helleday, T. (2009). Transcription-associated recombination in eukaryotes: link between transcription, replication and recombination. Mutagenesis *24*, 203-210.

Gould, G.M., Paggi, J.M., Guo, Y., Phizicky, D.V., Zinshteyn, B., Wang, E.T., Gilbert, W.V., Gifford, D.K., and Burge, C.B. (2016). Identification of new branch points and unconventional introns in Saccharomyces cerevisiae. Rna *22*, 1522-1534.

Green, P., Ewing, B., Miller, W., Thomas, P.J., and Green, E.D. (2003). Transcription-associated mutational asymmetry in mammalian evolution. Nature genetics *33*, 514-517.

Gressel, S., Schwalb, B., and Cramer, P. (2019). The pause-initiation limit restricts transcription activation in human cells. Nature communications *10*, 3603.

Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife *6*.

Grosso, A.R., Leite, A.P., Carvalho, S., Matos, M.R., Martins, F.B., Vitor, 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*.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Halasz, L., Karanyi, Z., Boros-Olah, B., Kuik-Rozsa, T., Sipos, E., Nagy, E., Mosolygo, L.A., Mazlo, A., Rajnavolgyi, E., Halmos, G.*, et al.* (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. Genome Res *27*, 1063-1073.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. Cell *170*, 774-786 e719.

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

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

Han, C., Khodadadi-Jamayran, A., Lorch, A.H., Jin, Q., Serafin, V., Zhu, P., Politanska, Y., Sun, L., Gutierrez-Diaz, B.T., Pryzhkova, M.V.*, et al.* (2022). SF3B1 homeostasis is critical for survival and therapeutic response in T cell leukemia. Sci Adv *8*, eabj8357.

Han, X., Yu, D., Gu, R., Jia, Y., Wang, Q., Jaganathan, A., Yang, X., Yu, M., Babault, N., Zhao, C.*, et al.* (2020). Roles of the BRD4 short isoform in phase separation and active gene transcription. Nat Struct Mol Biol *27*, 333-341.

Hantsche, M., and Cramer, P. (2016). The Structural Basis of Transcription: 10 Years After the Nobel Prize in Chemistry. Angew Chem Int Ed Engl *55*, 15972-15981.

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. Nature reviews Molecular cell biology *18*, 263-273.

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.

Hartono, S.R., Korf, I.F., and Chedin, F. (2015). GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. Nucleic acids research *43*, 9729-9741.

Hartono, S.R., Malapert, A., Legros, P., Bernard, P., Chedin, F., and Vanoosthuyse, V. (2018). The Affinity of the S9.6 Antibody for Double-Stranded RNAs Impacts the Accurate Mapping of R-Loops in Fission Yeast. J Mol Biol *430*, 272-284.

Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L*., et al.* (2018). Transcription Elongation Can Affect Genome 3D Structure. Cell *174*, 1522-1536 e1522.

Hendrix, D.A., Hong, J.W., Zeitlinger, J., Rokhsar, D.S., and Levine, M.S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc Natl Acad Sci U S A *105*, 7762-7767.

Hennig, T., Michalski, M., Rutkowski, A.J., Djakovic, L., Whisnant, A.W., Friedl, M.S., Jha, B.A., Baptista, M.A.P., L'Hernault, A., Erhard, F*., et al.* (2018). HSV-1-induced disruption of transcription termination resembles a cellular stress response but selectively increases chromatin accessibility downstream of genes. PLoS pathogens *14*, e1006954.

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. Molecular cell *52*, 517-528.

Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev *32*, 26-41.

Herold, S., Kalb, J., Buchel, G., Ade, C.P., Baluapuri, A., Xu, J., Koster, J., Solvie, D., Carstensen, A., Klotz, C*., et al.* (2019). Recruitment of BRCA1 limits MYCN-driven accumulation of stalled RNA polymerase. Nature *567*, 545-549.

Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. Nature reviews Molecular cell biology *18*, 637-650.

Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E.T., Nijmeijer, B., Kotthoff, C., Rybin, V., Amlacher, S*., et al.* (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. Nature *499*, 111-114.

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

Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J.Y., Yehia, G., and Tian, B. (2013). Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. Nature methods *10*, 133-139.

Howe, K.J., Kane, C.M., and Ares, M., Jr. (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. Rna *9*, 993-1006.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev *26*, 2119-2137.

Huang, F.T., Yu, K., Balter, B.B., Selsing, E., Oruc, Z., Khamlichi, A.A., Hsieh, C.L., and Lieber, M.R. (2007). Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. Molecular and cellular biology *27*, 5921-5932.

Huang, F.T., Yu, K., Hsieh, C.L., and Lieber, M.R. (2006). Downstream boundary of chromosomal R-loops at murine switch regions: implications for the mechanism of class switch recombination. Proc Natl Acad Sci U S A *103*, 5030-5035.

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

Huppert, J.L. (2008). Thermodynamic prediction of RNA-DNA duplex-forming regions in the human genome. Mol Biosyst *4*, 686-691.

Iacovoni, J.S., Caron, P., Lassadi, I., Nicolas, E., Massip, L., Trouche, D., and Legube, G. (2010). High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. The EMBO journal *29*, 1446-1457.

Imhof, A., Yang, X.J., Ogryzko, V.V., Nakatani, Y., Wolffe, A.P., and Ge, H. (1997). Acetylation of general transcription factors by histone acetyltransferases. Curr Biol *7*, 689-692.

Ishibashi, T., Dangkulwanich, M., Coello, Y., Lionberger, T.A., Lubkowska, L., Ponticelli, A.S., Kashlev, M., and Bustamante, C. (2014). Transcription factors IIS and IIF enhance transcription efficiency by differentially modifying RNA polymerase pausing dynamics. Proc Natl Acad Sci U S A *111*, 3419-3424.

Itoh, T., and Tomizawa, J. (1980). Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H. Proc Natl Acad Sci U S A *77*, 2450-2454.

Izban, M.G., and Luse, D.S. (1992). The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. Genes Dev *6*, 1342-1356.

Jacobs, E.Y., Ogiwara, I., and Weiner, A.M. (2004). Role of the C-terminal domain of RNA polymerase II in U2 snRNA transcription and 3' processing. Molecular and cellular biology *24*, 846-855.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.S., Brady, J.N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. Molecular cell *19*, 523-534.

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.

Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, B.T., Gnatt, A., Ren, B., and Roeder, R.G. (2012). Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIF with Pol II. Molecular cell *45*, 51-63.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nature reviews Molecular cell biology *16*, 167-177.

Ju, B.G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2006). A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription. Science *312*, 1798-1802.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kaida, D., Berg, M.G., Younis, I., Kasim, M., Singh, L.N., Wan, L., and Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature *468*, 664-668.

Kaneko, S., Chu, C., Shatkin, A.J., and Manley, J.L. (2007). Human capping enzyme promotes formation of transcriptional R loops in vitro. Proc Natl Acad Sci U S A *104*, 17620-17625.

Kashyap, M.K., Kumar, D., Villa, R., La Clair, J.J., Benner, C., Sasik, R., Jones, H., Ghia, E.M., Rassenti, L.Z., Kipps, T.J.*, et al.* (2015). Targeting the spliceosome in chronic lymphocytic leukemia with the macrolides FD-895 and pladienolide-B. Haematologica *100*, 945-954.

Katahira, J., Inoue, H., Hurt, E., and Yoneda, Y. (2009). Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. EMBO J *28*, 556-567.

Khodor, Y.L., Rodriguez, J., Abruzzi, K.C., Tang, C.H., Marr, M.T., 2nd, and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. Genes Dev *25*, 2502-2512.

Kim Guisbert, K.S., and Guisbert, E. (2017). SF3B1 is a stress-sensitive splicing factor that regulates both HSF1 concentration and activity. PloS one *12*, e0176382.

Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. Nature reviews Genetics *13*, 204-214.

Kireeva, M.L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M., and Studitsky, V.M. (2002). Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. Molecular cell *9*, 541-552.

Koga, M., Hayashi, M., and Kaida, D. (2015). Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. Nucleic acids research *43*, 8258-8267.

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

Kogoma, T., Skarstad, K., Boye, E., von Meyenburg, K., and Steen, H.B. (1985). RecA protein acts at the initiation of stable DNA replication in rnh mutants of Escherichia coli K-12. J Bacteriol *163*, 439-444.

Kogoma, T., and von Meyenburg, K. (1983). The origin of replication, oriC, and the dnaA protein are dispensable in stable DNA replication (sdrA) mutants of Escherichia coli K-12. EMBO J *2*, 463-468.

Kohler, A., and Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. Nature reviews Molecular cell biology *8*, 761-773.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat Chem Biol *3*, 570-575.

Kouzine, F., and Levens, D. (2007). Supercoil-driven DNA structures regulate genetic transactions. Front Biosci *12*, 4409-4423.

Kouzine, F., Liu, J., Sanford, S., Chung, H.J., and Levens, D. (2004). The dynamic response of upstream DNA to transcription-generated torsional stress. Nat Struct Mol Biol *11*, 1092-1100.

Kouzminova, E.A., and Kuzminov, A. (2021). Ultraviolet-induced RNA:DNA hybrids interfere with chromosomal DNA synthesis. Nucleic acids research *49*, 3888-3906.

Kraemer, S.M., Ranallo, R.T., Ogg, R.C., and Stargell, L.A. (2001). TFIIA interacts with TFIID via association with TATA-binding protein and TAF40. Molecular and cellular biology *21*, 1737-1746.

Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schubeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Molecular cell *67*, 411-422 e414.

Kreuzer, K.N., and Brister, J.R. (2010). Initiation of bacteriophage T4 DNA replication and replication fork dynamics: a review in the Virology Journal series on bacteriophage T4 and its relatives. Virology journal *7*, 358.

Kuhlman, T.C., Cho, H., Reinberg, D., and Hernandez, N. (1999). The general transcription factors IIA, IIB, IIF, and IIE are required for RNA polymerase II transcription from the human U1 small nuclear RNA promoter. Molecular and cellular biology *19*, 2130-2141.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950-953.

Kwak, H., and Lis, J.T. (2013). Control of transcriptional elongation. Annual review of genetics *47*, 483-508.

Kyburz, A., Friedlein, A., Langen, H., and Keller, W. (2006). Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing. Molecular cell *23*, 195-205.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nature reviews Molecular cell biology *18*, 548-562.

Lam, F.C., Kong, Y.W., Huang, Q., Vu Han, T.L., Maffa, A.D., Kasper, E.M., and Yaffe, M.B. (2020). BRD4 prevents the accumulation of R-loops and protects against transcription-replication collision events and DNA damage. Nature communications *11*, 4083.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans *34*, 1062-1066.

Landsverk, H.B., Sandquist, L.E., Sridhara, S.C., Rodland, G.E., Sabino, J.C., de Almeida, S.F., Grallert, B., Trinkle-Mulcahy, L., and Syljuasen, R.G. (2019). Regulation of ATR activity via the RNA polymerase II associated factors CDC73 and PNUTS-PP1. Nucleic acids research *47*, 1797-1813.

Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., and Merrikh, H. (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. Cell *170*, 787-799 e718.

Lee, D.Y., and Clayton, D.A. (1996). Properties of a primer RNA-DNA hybrid at the mouse mitochondrial DNA leading-strand origin of replication. J Biol Chem *271*, 24262-24269.

Lee, D.Y., and Clayton, D.A. (1998). Initiation of mitochondrial DNA replication by transcription and R-loop processing. J Biol Chem *273*, 30614-30621.

Lee, J.T. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nature reviews Molecular cell biology *12*, 815-826.

Legros, P., Malapert, A., Niinuma, S., Bernard, P., and Vanoosthuyse, V. (2014). RNA processing factors Swd2.2 and Sen1 antagonize RNA Pol III-dependent transcription and the localization of condensin at Pol III genes. PLoS genetics *10*, e1004794.

Leng, X., Ivanov, M., Kindgren, P., Malik, I., Thieffry, A., Brodersen, P., Sandelin, A., Kaplan, C.D., and Marquardt, S. (2020). Organismal benefits of transcription speed control at gene boundaries. EMBO reports *21*, e49315.

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

Li, X., and Manley, J.L. (2006). Cotranscriptional processes and their influence on genome stability. Genes Dev *20*, 1838-1847.

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

Lim, L.P., and Burge, C.B. (2001). A computational analysis of sequence features involved in recognition of short introns. Proc Natl Acad Sci U S A *98*, 11193-11198.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709-715.

Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry *11*, 3610-3618.

Lis, J. (1998). Promoter-associated pausing in promoter architecture and postinitiation transcriptional regulation. Cold Spring Harb Symp Quant Biol *63*, 347-356.

Lis, J.T., Mason, P., Peng, J., Price, D.H., and Werner, J. (2000). P-TEFb kinase recruitment and function at heat shock loci. Genes Dev *14*, 792-803.

Liu, H.L., and Cheng, S.C. (2012). The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. Molecular and cellular biology *32*, 5056-5066.

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, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A., and Rosenfeld, M.G. (2013). Brd4 and JMJD6-associated anti-pause enhancers in regulation of transcriptional pause release. Cell *155*, 1581-1595.

Liu, X., Freitas, J., Zheng, D., Oliveira, M.S., Hoque, M., Martins, T., Henriques, T., Tian, B., and Moreira, A. (2017). Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in Drosophila melanogaster. Rna *23*, 1807-1816.

Lu, H., Zawel, L., Fisher, L., Egly, J.M., and Reinberg, D. (1992). Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. Nature *358*, 641-645.

Luna, R., Rondon, A.G., Perez-Calero, C., Salas-Armenteros, I., and Aguilera, A. (2019). The THO Complex as a Paradigm for the Prevention of Cotranscriptional R-Loops. Cold Spring Harb Symp Quant Biol *84*, 105-114.

Ma, J., Bai, L., and Wang, M.D. (2013). Transcription under torsion. Science *340*, 1580-1583.

Ma, J., and Wang, M.D. (2016). DNA supercoiling during transcription. Biophys Rev *8*, 75-87.

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

Maduike, N.Z., Tehranchi, A.K., Wang, J.D., and Kreuzer, K.N. (2014). Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. Mol Microbiol *91*, 39-56.

Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G., and Lis, J.T. (2016). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. Molecular cell *62*, 63-78.

Malig, M., Hartono, S., Giafaglione, J., Sanz, L.A., and Chedin, F. (2020a). Ultra-Deep Coverage Single-Molecule R-loop Footprinting Reveals Principles of R-loop Formation J Mol Biol *in press*.

Malig, M., Hartono, S.R., Giafaglione, J.M., Sanz, L.A., and Chedin, F. (2020b). Ultra-deep Coverage Single-molecule R-loop Footprinting Reveals Principles of R-loop Formation. J Mol Biol *432*, 2271-2288.

Martins, S.B., Rino, J., Carvalho, T., Carvalho, C., Yoshida, M., Klose, J.M., de Almeida, S.F., and Carmo-Fonseca, M. (2011). Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. Nat Struct Mol Biol *18*, 1115-1123.

Masse, E., Phoenix, P., and Drolet, M. (1997). DNA topoisomerases regulate R-loop formation during transcription of the rrnB operon in Escherichia coli. J Biol Chem *272*, 12816-12823.

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

Masukata, H., and Tomizawa, J. (1990). A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. Cell *62*, 331-338.

Maxon, M.E., Goodrich, J.A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. Genes Dev *8*, 515-524.

Mayer, A., Landry, H.M., and Churchman, L.S. (2017). Pause & go: from the discovery of RNA polymerase pausing to its functional implications. Current opinion in cell biology *46*, 72-80.

Mayr, C. (2019). What Are 3' UTRs Doing? Cold Spring Harb Perspect Biol *11*.

McDaniel, S.L., and Strahl, B.D. (2017). Shaping the cellular landscape with Set2/SETD2 methylation. Cell Mol Life Sci *74*, 3317-3334.

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

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. Genes Dev *19*, 1401-1415.

Mercer, T.R., Clark, M.B., Andersen, S.B., Brunck, M.E., Haerty, W., Crawford, J., Taft, R.J., Nielsen, L.K., Dinger, M.E., and Mattick, J.S. (2015). Genome-wide discovery of human splicing branchpoints. Genome Res *25*, 290-303.

Metkar, M., Ozadam, H., Lajoie, B.R., Imakaev, M., Mirny, L.A., Dekker, J., and Moore, M.J. (2018). Higher-Order Organization Principles of Pre-translational mRNPs. Molecular cell *72*, 715-726 e713.

Mizui, Y., Sakai, T., Iwata, M., Uenaka, T., Okamoto, K., Shimizu, H., Yamori, T., Yoshimatsu, K., and Asada, M. (2004). Pladienolides, new substances from culture of Streptomyces platensis Mer-11107. III. In vitro and in vivo antitumor activities. The Journal of antibiotics *57*, 188-196.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688-700.

Morales, J.C., Richard, P., Patidar, P.L., Motea, E.A., Dang, T.T., Manley, J.L., and Boothman, D.A. (2016). XRN2 Links Transcription Termination to DNA Damage and Replication Stress. PLoS genetics *12*, e1006107.

Morimoto, S., Tsuda, M., Bunch, H., Sasanuma, H., Austin, C., and Takeda, S. (2019). Type II DNA Topoisomerases Cause Spontaneous Double-Strand Breaks in Genomic DNA. Genes (Basel) *10*.

Muhar, M., Ebert, A., Neumann, T., Umkehrer, C., Jude, J., Wieshofer, C., Rescheneder, P., Lipp, J.J., Herzog, V.A., Reichholf, B.*, et al.* (2018). SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science *360*, 800-805.

Mukhopadhyay, R., Lajugie, J., Fourel, N., Selzer, A., Schizas, M., Bartholdy, B., Mar, J., Lin, C.M., Martin, M.M., Ryan, M.*, et al.* (2014). Allele-specific genome-wide profiling in human primary erythroblasts reveal replication program organization. PLoS genetics *10*, e1004319.

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

Muniz, L., Deb, M.K., Aguirrebengoa, M., Lazorthes, S., Trouche, D., and Nicolas, E. (2017). Control of Gene Expression in Senescence through Transcriptional Read-Through of Convergent Protein-Coding Genes. Cell Rep *21*, 2433-2446.

Muniz, L., Nicolas, E., and Trouche, D. (2021). RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. EMBO J *40*, e105740.

Mutreja, K., Krietsch, J., Hess, J., Ursich, S., Berti, M., Roessler, F.K., Zellweger, R., Patra, M., Gasser, G., and Lopes, M. (2018). ATR-Mediated Global Fork Slowing and Reversal Assist Fork Traverse and Prevent Chromosomal Breakage at DNA Interstrand Cross-Links. Cell Rep *24*, 2629-2642 e2625.

Negrini, S., Gorgoulis, V.G., and Halazonetis, T.D. (2010). Genomic instability--an evolving hallmark of cancer. Nature reviews Molecular cell biology *11*, 220-228.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. Nature *543*, 72-77.

Nguyen, H.D., Leong, W.Y., Li, W., Reddy, P.N.G., Sullivan, J.D., Walter, M.J., Zou, L., and Graubert, T.A. (2018). Spliceosome Mutations Induce R Loop-Associated Sensitivity to ATR Inhibition in Myelodysplastic Syndromes. Cancer research *78*, 5363-5374.

Nguyen, H.D., Yadav, T., Giri, S., Saez, B., Graubert, T.A., and Zou, L. (2017). Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. Molecular cell *65*, 832-847 e834.

Nguyen, H.D., Zou, L., and Graubert, T.A. (2019). Targeting R-loop-associated ATR response in myelodysplastic syndrome. Oncotarget *10*, 2581-2582.

Nguyen, T.H., Galej, W.P., Bai, X.C., Savva, C.G., Newman, A.J., Scheres, S.H., and Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. Nature *523*, 47-52.

Nguyen, T.H.D., Galej, W.P., Bai, X.C., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 A resolution. Nature *530*, 298-302.

Niehrs, C., and Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. Nat Rev Mol Cell Biol *21*, 167-178.

Nikolov, D.B., and Burley, S.K. (1997). RNA polymerase II transcription initiation: a structural view. Proc Natl Acad Sci U S A *94*, 15-22.

Noe Gonzalez, M., Blears, D., and Svejstrup, J.Q. (2021). Causes and consequences of RNA polymerase II stalling during transcript elongation. Nature reviews Molecular cell biology *22*, 3-21.

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.

Nudler, E. (2012). RNA polymerase backtracking in gene regulation and genome instability. Cell *149*, 1438-1445.

Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell *89*, 33-41.

O'Brien, K., Matlin, A.J., Lowell, A.M., and Moore, M.J. (2008). The biflavonoid isoginkgetin is a general inhibitor of Pre-mRNA splicing. J Biol Chem *283*, 33147-33154.

Oesterreich, F.C., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell *165*, 372-381.

Ogawa, T., Pickett, G.G., Kogoma, T., and Kornberg, A. (1984). RNase H confers specificity in the dnaA-dependent initiation of replication at the unique origin of the Escherichia coli chromosome in vivo and in vitro. Proc Natl Acad Sci U S A *81*, 1040-1044.

Oh, J.M., Di, C., Venters, C.C., Guo, J., Arai, C., So, B.R., Pinto, A.M., Zhang, Z., Wan, L., Younis, I.*, et al.* (2017). U1 snRNP telescripting regulates a size-function-stratified human genome. Nat Struct Mol Biol *24*, 993-999.

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 *167*, 1001-1013 e1007.

Ohrt, T., Odenwalder, P., Dannenberg, J., Prior, M., Warkocki, Z., Schmitzova, J., Karaduman, R., Gregor, I., Enderlein, J., Fabrizio, P.*, et al.* (2013). Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. Rna *19*, 902-915.

Onyango, D.O., Lee, G., and Stark, J.M. (2017). PRPF8 is important for BRCA1-mediated homologous recombination. Oncotarget *8*, 93319-93337.

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. Molecular cell *35*, 228-239.

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z.P.*, et al.* (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. Cell *161*, 774-789.

Petesch, S.J., and Lis, J.T. (2012). Overcoming the nucleosome barrier during transcript elongation. Trends in genetics : TIG *28*, 285-294.

Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thermes, C., Chen, C.L., and Hyrien, O. (2016). Replication landscape of the human genome. Nature communications *7*, 10208.

Phillips, D.D., Garboczi, D.N., Singh, K., Hu, Z., Leppla, S.H., and Leysath, C.E. (2013). The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. J Mol Recognit *26*, 376-381.

Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y.*, et al.* (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281-1295.

Picard, F., Cadoret, J.C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L., and Prioleau, M.N. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS genetics *10*, e1004282.

Pinto, P.A., Henriques, T., Freitas, M.O., Martins, T., Domingues, R.G., Wyrzykowska, P.S., Coelho, P.A., Carmo, A.M., Sunkel, C.E., Proudfoot, N.J.*, et al.* (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. EMBO J *30*, 2431-2444.

Polak, P., and Arndt, P.F. (2008). Transcription induces strand-specific mutations at the 5' end of human genes. Genome Res *18*, 1216-1223.

Pommier, Y., Sun, Y., Huang, S.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. Nature reviews Molecular cell biology *17*, 703-721.

Promonet, A., Padioleau, I., Liu, Y., Sanz, L., Biernacka, A., Schmitz, A.L., Skrzypczak, M., Sarrazin, A., Mettling, C., Rowicka, M.*, et al.* (2020). Topoisomerase 1 prevents replication stress at R-loop-enriched transcription termination sites. Nature communications *11*, 3940.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Qin, D., Huang, L., Wlodaver, A., Andrade, J., and Staley, J.P. (2016). Sequencing of lariat termini in S. cerevisiae reveals 5' splice sites, branch points, and novel splicing events. Rna *22*, 237-253.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Ramachandran, S., Tran, D.D., Klebba-Faerber, S., Kardinal, C., Whetton, A.D., and Tamura, T. (2011). An ataxia-telangiectasia-mutated (ATM) kinase mediated response to DNA damage down-regulates the mRNA-binding potential of THOC5. Rna *17*, 1957-1966.

Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.T., Urlaub, H., Kastner, B.*, et al.* (2016). Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science *353*, 1399-1405.

Reaban, M.E., Lebowitz, J., and Griffin, J.A. (1994). Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. J Biol Chem *269*, 21850-21857.

Reed, R., and Cheng, H. (2005). TREX, SR proteins and export of mRNA. Current opinion in cell biology *17*, 269-273.

Reed, R., and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. Cell *108*, 523-531.

Reed, R., and Maniatis, T. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. Cell *41*, 95-105.

Reimer, K.A., Mimoso, C.A., Adelman, K., and Neugebauer, K.M. (2021). Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. Molecular cell *81*, 998-1012 e1017.

Richard, P., and Manley, J.L. (2017). R Loops and Links to Human Disease. J Mol Biol *429*, 3168-3180.

Rigo, F., and Martinson, H.G. (2008). Functional coupling of last-intron splicing and 3'-end processing to transcription in vitro: the poly(A) signal couples to splicing before committing to cleavage. Molecular and cellular biology *28*, 849-862.

Rigo, F., and Martinson, H.G. (2009). Polyadenylation releases mRNA from RNA polymerase II in a process that is licensed by splicing. Rna *15*, 823-836.

Rigo, N., Sun, C., Fabrizio, P., Kastner, B., and Luhrmann, R. (2015). Protein localisation by electron microscopy reveals the architecture of the yeast spliceosomal B complex. EMBO J *34*, 3059-3073.

Rimel, J.K., and Taatjes, D.J. (2018). The essential and multifunctional TFIIH complex. Protein Sci *27*, 1018-1037.

Rollins, R.A., Korom, M., Aulner, N., Martens, A., and Dorsett, D. (2004). Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. Molecular and cellular biology *24*, 3100-3111.

Rosa-Mercado, N.A., Zimmer, J.T., Apostolidi, M., Rinehart, J., Simon, M.D., and Steitz, J.A. (2021). Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. Molecular cell *81*, 502-513 e504.

Rougvie, A.E., and Lis, J.T. (1988). The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. Cell *54*, 795-804.

Roy, D., Zhang, Z., Lu, Z., Hsieh, C.L., and Lieber, M.R. (2010). Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. Molecular and cellular biology *30*, 146-159.

Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. Nature reviews Molecular cell biology *16*, 129-143.

Sakharkar, M.K., Perumal, B.S., Sakharkar, K.R., and Kangueane, P. (2005). An analysis on gene architecture in human and mouse genomes. In Silico Biol *5*, 347-365.

Saldi, T., Fong, N., and Bentley, D.L. (2018). Transcription elongation rate affects nascent histone pre-mRNA folding and 3' end processing. Genes Dev *32*, 297-308.

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

Sanz, L.A., and Chedin, F. (2019a). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc *14*, 1734-1755.

Sanz, L.A., and Chedin, F. (2019b). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. Nat Protoc.

Sanz, L.A., Hartono, S.R., Lim, Y.W., Steyaert, S., Rajpurkar, A., Ginno, P.A., Xu, X., and Chedin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. Molecular cell *63*, 167-178.

Saran, S., Tran, D.D., Ewald, F., Koch, A., Hoffmann, A., Koch, M., Nashan, B., and Tamura, T. (2016). Depletion of three combined THOC5 mRNA export protein target genes synergistically induces human hepatocellular carcinoma cell death. Oncogene *35*, 3872-3879.

Saunders, A., Werner, J., Andrulis, E.D., Nakayama, T., Hirose, S., Reinberg, D., and Lis, J.T. (2003). Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. Science *301*, 1094-1096.

Schilbach, S., Aibara, S., Dienemann, C., Grabbe, F., and Cramer, P. (2021). Structure of RNA polymerase II pre-initiation complex at 2.9 A defines initial DNA opening. Cell *184*, 4064-4072 e4028.

Schneider, C., and Tollervey, D. (2013). Threading the barrel of the RNA exosome. Trends Biochem Sci *38*, 485-493.

Schroder, S., Cho, S., Zeng, L., Zhang, Q., Kaehlcke, K., Mak, L., Lau, J., Bisgrove, D., Schnolzer, M., Verdin, E.*, et al.* (2012). Two-pronged binding with bromodomain-containing protein 4 liberates positive transcription elongation factor b from inactive ribonucleoprotein complexes. J Biol Chem *287*, 1090-1099.

Schwer, B. (2008). A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Molecular cell *30*, 743-754.

Seiler, J.A., Conti, C., Syed, A., Aladjem, M.I., and Pommier, Y. (2007). The intra-S-phase checkpoint affects both DNA replication initiation and elongation: single-cell and -DNA fiber analyses. Molecular and cellular biology *27*, 5806-5818.

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS genetics *5*, e1000446.

Serfling, E., Jasin, M., and Schaffner, W. (1985). Enhancers and eukaryotic gene transcription. Trends in Genetics *1*, Pages 224-230.

Shalgi, R., Hurt, J.A., Lindquist, S., and Burge, C.B. (2014). Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. Cell Rep *7*, 1362-1370.

Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nature genetics *49*, 1045-1051.

Sheridan, R.M., Fong, N., D'Alessandro, A., and Bentley, D.L. (2019). Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5' Pause Release, Termination, and Transcription Elongation Rate. Molecular cell *73*, 107-118 e104.

Singh, G., Pratt, G., Yeo, G.W., and Moore, M.J. (2015). The Clothes Make the mRNA: Past and Present Trends in mRNP Fashion. Annual review of biochemistry *84*, 325-354.

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. Molecular cell *42*, 794-805.

Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A., and Pombo, A. (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. Molecular cell *73*, 930-945 e934.

Smolka, J., Hartono, S.R., Sanz, L.A., and Chedin, F. (2020). Recognition of cellular RNAs by the S9.6 antibody creates pervasive artefacts when imaging RNA:DNA hybrids.

Smolka, J.A., Sanz, L.A., Hartono, S.R., and Chedin, F. (2021a). Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. The Journal of cell biology *220*.

Smolka, J.A., Sanz, L.A., Hartono, S.R., and Chedin, F. (2021b). Recognition of RNAs by the S9.6 antibody creates pervasive artefacts when imaging RNA:DNA hybrids. Journal of Cell Biology *in press*.

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

Sollier, J., Stork, C.T., Garcia-Rubio, M.L., Paulsen, R.D., Aguilera, A., and Cimprich, K.A. (2014). Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. Molecular cell *56*, 777-785.

Sousa-Luis, R., Dujardin, G., Zukher, I., Kimura, H., Weldon, C., Carmo-Fonseca, M., Proudfoot, N.J., and Nojima, T. (2021). POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. Molecular cell *81*, 1935-1950 e1936.

Sridhara, S.C., Carvalho, S., Grosso, A.R., Gallego-Paez, L.M., Carmo-Fonseca, M., and de Almeida, S.F. (2017). Transcription Dynamics Prevent RNA-Mediated Genomic Instability through SRPK2-Dependent DDX23 Phosphorylation. Cell Rep *18*, 334-343.

St Germain, C.P., Zhao, H., Sinha, V., Sanz, L., Chedin, F., and Barlow, J.H. (2021). Genomic patterns of transcription-replication interactions in mouse primary B cells. biorXiv.

Stirling, P.C., Chan, Y.A., Minaker, S.W., Aristizabal, M.J., Barrett, I., Sipahimalani, P., Kobor, M.S., and Hieter, P. (2012). R-loop-mediated genome instability in mRNA cleavage and polyadenylation mutants. Genes Dev *26*, 163-175.

Stolz, R., Sulthana, S., Hartono, S.R., Malig, M., Benham, C.J., and Chedin, F. (2019a). Interplay between DNA sequence and negative superhelicity drives R-loop structures. Proc Natl Acad Sci U S A *116*, 6260-6269.

Stolz, R., Sulthana, S., Hartono, S.R., Malig, M., Benham, C.J., and Chedin, F. (2019b). Interplay between DNA sequence and negative superhelicity drives R-loop structures. Proc Natl Acad Sci U S A.

Stork, C.T., Bocek, M., Crossley, M.P., Sollier, J., Sanz, L.A., Chedin, F., Swigut, T., and Cimprich, K.A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. Elife *5*.

Stuckey, R., Garcia-Rodriguez, N., Aguilera, A., and Wellinger, R.E. (2015). Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. Proc Natl Acad Sci U S A *112*, 5779-5784.

Sun, C. (2020). The SF3b complex: splicing and beyond. Cell Mol Life Sci.

Svikovic, S., Crisp, A., Tan-Wong, S.M., Guilliam, T.A., Doherty, A.J., Proudfoot, N.J., Guilbaud, G., and Sale, J.E. (2019). R-loop formation during S phase is restricted by PrimPol-mediated repriming. EMBO J *38*.

Taggart, A.J., Lin, C.L., Shrestha, B., Heintzelman, C., Kim, S., and Fairbrother, W.G. (2017). Large-scale analysis of branchpoint usage across species and cell lines. Genome Res *27*, 639-649.

Takahashi, H., Parmely, T.J., Sato, S., Tomomori-Sato, C., Banks, C.A., Kong, S.E., Szutorisz, H., Swanson, S.K., Martin-Brown, S., Washburn, M.P.*, et al.* (2011). Human mediator subunit MED26 functions as a docking site for transcription elongation factors. Cell *146*, 92-104.

Tan-Wong, S.M., Dhir, S., and Proudfoot, N.J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. Mol Cell *76*, 600-616 e606.

Teloni, F., Michelena, J., Lezaja, A., Kilic, S., Ambrosi, C., Menon, S., Dobrovolna, J., Imhof, R., Janscak, P., Baubec, T.*, et al.* (2019). Efficient Pre-mRNA Cleavage Prevents Replication-Stress-Associated Genome Instability. Molecular cell *73*, 670-683 e612.

Teves, S.S., Weber, C.M., and Henikoff, S. (2014). Transcribing through the nucleosome. Trends Biochem Sci *39*, 577-586.

Tran, D.D., Saran, S., Koch, A., and Tamura, T. (2016). mRNA export protein THOC5 as a tool for identification of target genes for cancer therapy. Cancer Lett *373*, 222-226.

Tresini, M., Warmerdam, D.O., Kolovos, P., Snijder, L., Vrouwe, M.G., Demmers, J.A., van, I.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.

Tuduri, S., Crabbe, L., Conti, C., Tourriere, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C*., et al.* (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. Nat Cell Biol *11*, 1315-1324.

Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. Molecular and cellular biology *26*, 9185-9195.

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D*., et al.* (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature *583*, 711-719.

Vanoosthuyse, V. (2018). Strengths and Weaknesses of the Current Strategies to Map and Characterize R-Loops. Noncoding RNA *4*.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nature reviews Molecular cell biology *16*, 178-189.

Verger, A., Monte, D., and Villeret, V. (2021). Take Your PIC. Trends Biochem Sci.

Vihervaara, A., Duarte, F.M., and Lis, J.T. (2018). Molecular mechanisms driving transcriptional stress responses. Nature reviews Genetics *19*, 385-397.

Vilborg, A., Passarelli, M.C., Yario, T.A., Tycowski, K.T., and Steitz, J.A. (2015). Widespread Inducible Transcription Downstream of Human Genes. Molecular cell *59*, 449-461.

Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., and Shalgi, R. (2017). Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proc Natl Acad Sci U S A *114*, E8362-E8371.

Vilborg, A., and Steitz, J.A. (2017). Readthrough transcription: How are DoGs made and what do they do? RNA biology *14*, 632-636.

Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II-DSIF-NELF. Nature *560*, 601-606.

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 Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016a). Structure of a yeast catalytic step I spliceosome at 3.4 A resolution. Science *353*, 895-904.

Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C., and Shi, Y. (2016b). The 3.8 A structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. Science *351*, 466-475.

Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M*., et al.* (2021). Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. Cell *184*, 2878-2895 e2820.

Wan, Y., Zheng, X., Chen, H., Guo, Y., Jiang, H., He, X., Zhu, X., and Zheng, Y. (2015). Splicing function of mitotic regulators links R-loop-mediated DNA damage to tumor cell killing. The Journal of cell biology *209*, 235-246.

Wang, J., Rojas, P., Mao, J., Muste Sadurni, M., Garnier, O., Xiao, S., Higgs, M.R., Garcia, P., and Saponaro, M. (2021a). Persistence of RNA transcription during DNA replication delays duplication of transcription start sites until G2/M. Cell Rep *34*, 108759.

Wang, K., Wang, H., Li, C., Yin, Z., Xiao, R., Li, Q., Xiang, Y., Wang, W., Huang, J., Chen, L.*, et al.* (2021b). Genomic profiling of native R loops with a DNA-RNA hybrid recognition sensor. Sci Adv *7*.

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

Williamson, L., Saponaro, M., Boeing, S., East, P., Mitter, R., Kantidakis, T., Kelly, G.P., Lobley, A., Walker, J., Spencer-Dene, B.*, et al.* (2017). UV Irradiation Induces a Non-coding RNA that Functionally Opposes the Protein Encoded by the Same Gene. Cell *168*, 843-855 e813.

Wimberly, H., Shee, C., Thornton, P.C., Sivaramakrishnan, P., Rosenberg, S.M., and Hastings, P.J. (2013). R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. Nature communications *4*, 2115.

Wissink, E.M., Vihervaara, A., Tippens, N.D., and Lis, J.T. (2019). Nascent RNA analyses: tracking transcription and its regulation. Nature reviews Genetics *20*, 705-723.

Wulfridge, P., and Sarma, K. (2021). A nuclease- and bisulfite-based strategy captures strand-specific R-loops genome-wide. Elife *10*.

Xu, B., and Clayton, D.A. (1995). A persistent RNA-DNA hybrid is formed during transcription at a phylogenetically conserved mitochondrial DNA sequence. Molecular and cellular biology *15*, 580-589.

Xu, W., Li, K., Li, S., Hou, Q., Zhang, Y., Liu, K., and Sun, Q. (2020). The R-Loop Atlas of Arabidopsis Development and Responses to Environmental Stimuli. Plant Cell *32*, 888-903.

Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., and Sun, Q. (2017). The R-loop is a common chromatin feature of the Arabidopsis genome. Nat Plants *3*, 704-714.

Yague-Sanz, C., Vanrobaeys, Y., Fernandez, R., Duval, M., Larochelle, M., Beaudoin, J., Berro, J., Labbe, S., Jacques, P.E., and Bachand, F. (2020). Nutrient-dependent control of RNA polymerase II elongation rate regulates specific gene expression programs by alternative polyadenylation. Genes Dev *34*, 883-897.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast activated spliceosome at 3.5 A resolution. Science *353*, 904-911.

Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017). Structure of a yeast step II catalytically activated spliceosome. Science *355*, 149-155.

Yan, Q., and Sarma, K. (2020). MapR: A Method for Identifying Native R-Loops Genome Wide. Curr Protoc Mol Biol *130*, e113.

Yan, Q., Shields, E.J., Bonasio, R., and Sarma, K. (2019). Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach. Cell Rep *29*, 1369-1380 e1365.

Yang, X., Liu, Q.L., Xu, W., Zhang, Y.C., Yang, Y., Ju, L.F., Chen, J., Chen, Y.S., Li, K., Ren, J.*, et al.* (2019). m(6)A promotes R-loop formation to facilitate transcription termination. Cell Res *29*, 1035-1038.

Yang, Z., He, N., and Zhou, Q. (2008). Brd4 recruits P-TEFb to chromosomes at late mitosis to promote G1 gene expression and cell cycle progression. Molecular and cellular biology *28*, 967-976.

Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K., and Zhou, Q. (2005). Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. Molecular cell *19*, 535-545.

Yokoi, A., Kotake, Y., Takahashi, K., Kadowaki, T., Matsumoto, Y., Minoshima, Y., Sugi, N.H., Sagane, K., Hamaguchi, M., Iwata, M.*, et al.* (2011). Biological validation that SF3b is a target of the antitumor macrolide pladienolide. FEBS J *278*, 4870-4880.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M.*, et al.* (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature *478*, 64-69.

Yost, H.J., and Lindquist, S. (1986). RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell *45*, 185-193.

Yu, K., Chedin, F., Hsieh, C.L., Wilson, T.E., and Lieber, M.R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nat Immunol *4*, 442-451.

Yu, K., and Lieber, M.R. (2019). Current insights into the mechanism of mammalian immunoglobulin class switch recombination. Crit Rev Biochem Mol Biol *54*, 333-351.

Zatreanu, D., Han, Z., Mitter, R., Tumini, E., Williams, H., Gregersen, L., Dirac-Svejstrup, A.B., Roma, S., Stewart, A., Aguilera, A.*, et al.* (2019). Elongation Factor TFIIS Prevents Transcription Stress and R-Loop Accumulation to Maintain Genome Stability. Molecular cell *76*, 57-69 e59.

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. Nature genetics *48*, 1385-1395.

Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. Nat Cell Biol *16*, 2-9.

Zhang, W., Prakash, C., Sum, C., Gong, Y., Li, Y., Kwok, J.J., Thiessen, N., Pettersson, S., Jones, S.J., Knapp, S.*, et al.* (2012). Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells. J Biol Chem *287*, 43137-43155.

Zhang, X., Chiang, H.C., Wang, Y., Zhang, C., Smith, S., Zhao, X., Nair, S.J., Michalek, J., Jatoi, I., Lautner, M.*, et al.* (2017). Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. Nature communications *8*, 15908.

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., and Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. Cell *138*, 1122-1136.