# Interrogating Putative Roles for R-Loops in dsRNA Formation and Transcription Regulation

Presented by

**Ahmed Said Abd-Elsattar**

A thesis submitted for the degree of Doctor of Philosophy (Genome Sciences)

11/2021

**Department of Genome Sciences**
**John Curtin School of Medical Research**
**College of Health and Medicine**

**The Australian National University**

# Interrogating Putative Roles for R-Loops in dsRNA Formation and Transcription Regulation

**Supervision panel**


**Assoc. Prof. Tamas Fischer** **Prof. Thomas Preiss**

**Prof. David Tremethick** **Prof. Leonie Quinn**

**Department of Genome Sciences**
**John Curtin School of Medical Research**
**College of Health and Medicine**

**The Australian National University**


**2021**

# Acknowledgment

I'd like to gain this chance and thank every single person who has provided any kind of help, guidance or support during the path of my PhD at JCSMR, ANU.

I'm very thankful for my primary supervisor, assoc. Prof. Tamas Fischer, for his support during the time of my PhD. He has given me the opportunity to work on many projects through which I tried and excelled multiple state-of-the-art techniques without which this work couldn't have been done.

I'd like to express my gratitude and gratefulness for Prof. Thomas Preiss, the co-head of Genome Sciences and Cancer Division, for his guidance during the writing stage of this work. He has been always open for help and discussion. His advice enabled me finalize the data analysis and writing part of this thesis and see the final picture of this work.

I'd like to also thank my co-supervisor, Prof. David Tremethick, the ex-head of Genome Sciences Department, for his support, fruitful discussions and helpful suggestions during the time of my PhD.

Great thanks to Dr. Attila Horvath for contributing to this work through conducting the bioinformatic analysis, without which this work couldn't have come to light.

I'd like to thank Fischer lab previous and current members, Dr. Geza Schermann, Dr. Jane Reid, Anusree Sivadas and Yoona Kim for their support and fellowship during the time of my PhD.

## Statement of originality

This is to certify that the content of this thesis is my own work and that any assistance received in preparing this thesis and sources produced by others have been acknowledged. This thesis has not been submitted for any other degree or purposes.

Ahmed S. Abd-Elsattar

## Authorship attribution statement

I have conducted all the experiments, generated the figures and written the manuscript.

Materials, data and analysis included in this thesis, but not obtained, generated or performed by me are listed below. They have been included in this thesis as they significantly contribute to the overall understanding and body of knowledge explored in this thesis. All authors affirm that the below are correct and have given consent to have their work included in this thesis.

T.F has generated heatmaps in Figure 4.5 for a previous ChIP-exo data from the lab.

A.H has conducted the bioinformatic analysis for all the sequencing data. He generated metaplots and heatmaps in Figures 4.14 and 4.15 and spearman correlation plots in Figures 4.16, 4.17, 4.24 and 4.25

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

| **Supervisor Name** | **Signature** | **Date** |
|---|---|---|

# Abstract

R-loops are three-stranded nucleic acid structures consisting of an RNA-DNA hybrid and a displaced ssDNA. Their formation is a conserved feature of the genomes of all living organisms, induced naturally by RNA transcription and specific nucleotide-sequence features. Despite being considered for long as transcriptional byproducts and threats to genomic integrity, R-loops are currently regarded as important regulatory structures required for multiple biological processes. Intriguingly, R-loops have been implicated in transcription termination and RNAi-mediated heterochromatin formation through dsRNA formation at specific gene terminators. Yet this role has never been confirmed or tested on a genome-wide scale, and the underlying mechanism is still unclear. Here, I proposed two models for R-loop-mediated dsRNA formation and examined them using *S. pombe* cells. The first model is based on bidirectional double R-loop formation while the second is based on R-loops coupled with overlapping free asRNA.

For a comprehensive and complementary R-loop profiling, I used three different S9.6-based and DRIP-like methods for consistent, high resolution and directional R-loop mapping. Beside the ultra-high-resolution double-stranded DNA-based ChIP-exonuclease (dsChIP-exo) technique, I developed directional and single stranded DNA-based ChIP-exonuclease (ssChIP-exo) to map the DNA strand of the hybrid. Moreover, I adopted the DRIPc method combined with the SMART-seq technology (SMART-DRIPc) to sequence the RNA strand of R-loops. To examine a possible role for R-loops in heterochromatin formation, I mapped the heterochromatic H3K9me2 mark using a high-sensitivity version of ChIP-exo and studied the impact of either RNase H depletion or overexpression on levels of this histone mark.

dsChIP-exo method revealed bidirectional signals mapping to both DNA strands, misleading to the belief that double R-loops form widely over the genome. Differently, both ssChIP-exo and SMART-DRIPc methods revealed single R-loops forming in transcription direction. Although my data pointed toward double R-loops formation over tRNAs and rRNAs, it revealed a global anti-correlation between template and non-template strand R-loops even over the double R-loop forming genes. I found that R-loops over 60 to 70% of forming regions were extremely sensitive to physiological RNase H levels, which weakens both models of R-loop-dependent dsRNA formation. However, a category of RNase H-insensitive R-loops were associated with dsRNA-forming sites of sense-antisense transcription. Unexpectedly, RNase

H depletion decreased R-loop signals over template strand of protein synthesis and ribosome biogenesis genes. R-loops of these genes are RNase H-insensitive and form on reverse strand. Excitingly, RNase H deletion, globally, increased forward-strand but decreased reverse-strand R-loop signals. Independently, no R-loops were detected over gene terminators or enriched over any of the heterochromatic repeats except tRNA genes and chromosome III telomeres rich in rRNA genes. Strikingly, either RNase H depletion or overexpression, respectively, disrupted H3K9me2 over heterochromatin in fission yeast and depleted global H3K9me2 in mammalian cells.

R-loops may have been overrated as contributors for transcription termination and heterochromatin formation as they stand as marks of transcriptionally-active rather than heterochromatic domains. Despite challenges for the currently proposed models, R-loops coupled with overlapping asRNAs seems to be a more plausible model for dsRNA formation compared to double R-loops model. Nevertheless, practical involvement of R-loops in dsRNA formation still needs to be biochemically confirmed. My results suggest that perturbing RNase H levels, independent from R-loops, may alter the transcriptome and proteome, and impact cellular activities. Interestingly, R-loop orientation seems to modulate its response to RNase H. Excitingly, my observations suggest a role for RNase H in regulation of gene expression through suppressing formation of forward-strand R-loops and maintaining expression of sense genes. Transcriptome profiling and quantitative single-cell R-loop mapping using mimic and spike-in R-loop standards are required to confirm these results.

**Keywords:** RNA-DNA hybrids, R-loops, DRIP, ChIP-exo, DRIPc, SMART, RNase H, asRNA, dsRNA, termination, heterochromatin, H3K9me2.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| **Ab** | Antibody |
| **AhR** | Aryl hydrocarbon receptor |
| **AID** | Activation-induced cytidine deaminase |
| **ARC** | Argonaute complex |
| **AS** | Antisense |
| **aslncRNA** | Antisense long non-coding RNA |
| **bisDRIP** | Bisulphite conversion after DRIP |
| **bp** | Base pair |
| **CBP** | Cycles per burst |
| **CD** | Central dichroism |
| **ChIP** | Chromatin immunoprecipitation |
| **ChIP-exo** | Chromatin immunoprecipitation exonuclease |
| **ChIP-exo-TT** | ChIP-exo terminal transferase |
| **ChIP-nexus** | Chromatin immunoprecipitation with nucleotide resolution through exonuclease, unique barcode and single ligation |
| **CLRC** | Cryptic loci regulator |
| **CoTC** | Co-transcriptional termination and cleavage |
| **CSR** | Class switch recombination |
| **CTD** | C-terminal domain |
| **CUT&RUN** | Cleavage under targets and release using nuclease |
| **CUT&Tag** | Cleavage under targets and tagmentation |
| **CUTs** | Cryptic unstable transcripts |
| **DF** | Duty factor |
| **DMSO** | Dimethyl-sulfoxide |
| **DR** | DNA purine-RNA pyrimidine |
| **DRIP** | DNA-RNA immunoprecipitation |
| **DRIPc** | DRIP followed by cDNA synthesis |
| **DRIVE** | DNA-RNA immunoprecipitation and *in vitro* enrichment |
| **DS** | Double stranded |
| **DSB** | Double strand break |

| | |
|---|---|
| **dsChIP-exo** | dsDNA-based ChIP-exo |
| **dsDNA** | Double stranded DNA |
| **dsRNA** | Double stranded RNA |
| **DTT** | Dithiothreitol |
| **ETS** | External transcribed spacer |
| **FA** | Formaldehyde |
| **FLC** | Flowering locus |
| **FTP** | FLAG TEV protein-A |
| **fw** | Forward |
| **GST** | Glutathione S-transferase |
| **GTF** | General transcription factor |
| **HBD** | Hybrid-binding domain |
| **HDAC** | Histone deacetylase complex |
| **HMT** | Histone methyltransferase |
| **HP1** | Heterochromatin protein 1 |
| **hsChIP-exo** | High-sensitivity ChIP-exo |
| **hsp** | Heat shock protein |
| **I** | Intensity |
| **IP** | Immunoprecipitation |
| **MBD** | Maltose binding domain |
| **min** | Minute |
| **MTS** | Mitochondrial targeting sequence |
| **NDR** | Nucleosome depleted region |
| **NMD** | Nonsense mediated decay |
| **no-XL** | Non-crosslinked |
| **nt** | Nucleotide |
| **NTC** | Non-template control |
| **OD** | Optical density |
| **ORF** | Open reading frame |
| **pA** | Polyadenylation |
| **PAP** | Poly-A polymerase |
| **PBS** | Phosphate buffered saline |

| | |
|---|---|
| **PcG** | Polycomb group |
| **PE** | Paired end |
| **PI** | Protease inhibitor |
| **PMSF** | Phenylmethylsulfonyl fluoride |
| **PNK** | Polynucleotide kinase |
| **Pol** | Polymerase |
| **PREs** | Polycomb Response Elements |
| **PROMPTs** | Promoter upstream transcripts |
| **RD** | RNA purine-DNA pyrimidine |
| **RDIP** | RNA-DNA immunoprecipitation |
| **RDRC** | RNA dependent RNA polymerase complex |
| **RDTT** | Rho dependent transcription termination |
| **REZ** | R-loop elongation zone |
| **RF** | Replication fork |
| **RITS** | RNA induced transcriptional silencing |
| **RIZ** | R-loop initiation zone |
| **RNAi** | RNA interference |
| **RNAP** | RNA polymerase |
| **RNH** | RNase H |
| **RNH OE** | RNase H overexpression |
| ***rnhΔΔ*** | *rnh1Δ rnh201Δ* |
| **RNIII** | RNase III |
| **RNP** | Ribonucleoprotein |
| **RPA** | Replication protein A |
| **rSAP** | Recombinant shrimp alkaline phosphatase |
| **RT** | Reverse transcription |
| **rv** | Reverse |
| **SDS** | Sodium dodecyl sulphate |
| **SE** | Single end |
| **sec** | Second |
| **siRNA** | Small interference RNA |
| **SL** | Splint ligation |

| | |
|---|---|
| **SMART** | Switch mechanism at 5’ end of RNA transcripts |
| **SMRF** | Single molecule R-loop foot printing |
| **SSBP** | Single stranded DNA binding protein |
| **ssChIP-exo** | ssDNA-based ChIP-exo |
| **ssDNA** | Single stranded DNA |
| **ssDRIP** | ssDNA-based DRIP |
| **SUTs** | Stable unannotated transcripts |
| **TBP** | TATA box-binding protein |
| **TCA** | Trichloroacetic acid |
| **TdT** | Terminal deoxy-transferase |
| **TEV** | Tobacco etch virus |
| **TF** | Transcription factor |
| **TRE** | Tet response elements |
| **TS** | Template switching |
| **TSP** | Template switching primer |
| **TSS** | Transcription start site |
| **tTS** | Tetracycline-controlled trans-suppressor system |
| **XL** | Crosslinked |
| **XUTs** | Xrn2 sensitive transcripts |

# Chapter 1. Introduction

## 1.1 Overview

R-loops are non-canonical three-stranded nucleic acid structures that from naturally over the genome of many organisms. They form mainly as a result of transcription when an RNA strand invades a DNA duplex of homologous sequence, thereby hybridizing to the template DNA strand and displacing the non-template one. R-loops have been associated with open chromatin and linked to pausing RNA polymerase. They form over purine-rich DNA sequences and have been correlated with G quadruplexes (Figure 1.1). In addition, R-loops can form co-transcriptionally (*in cis*) behind elongating RNA Pol, or post-transcriptionally (*in trans*) when the RNA invades a duplex DNA away from transcription site.

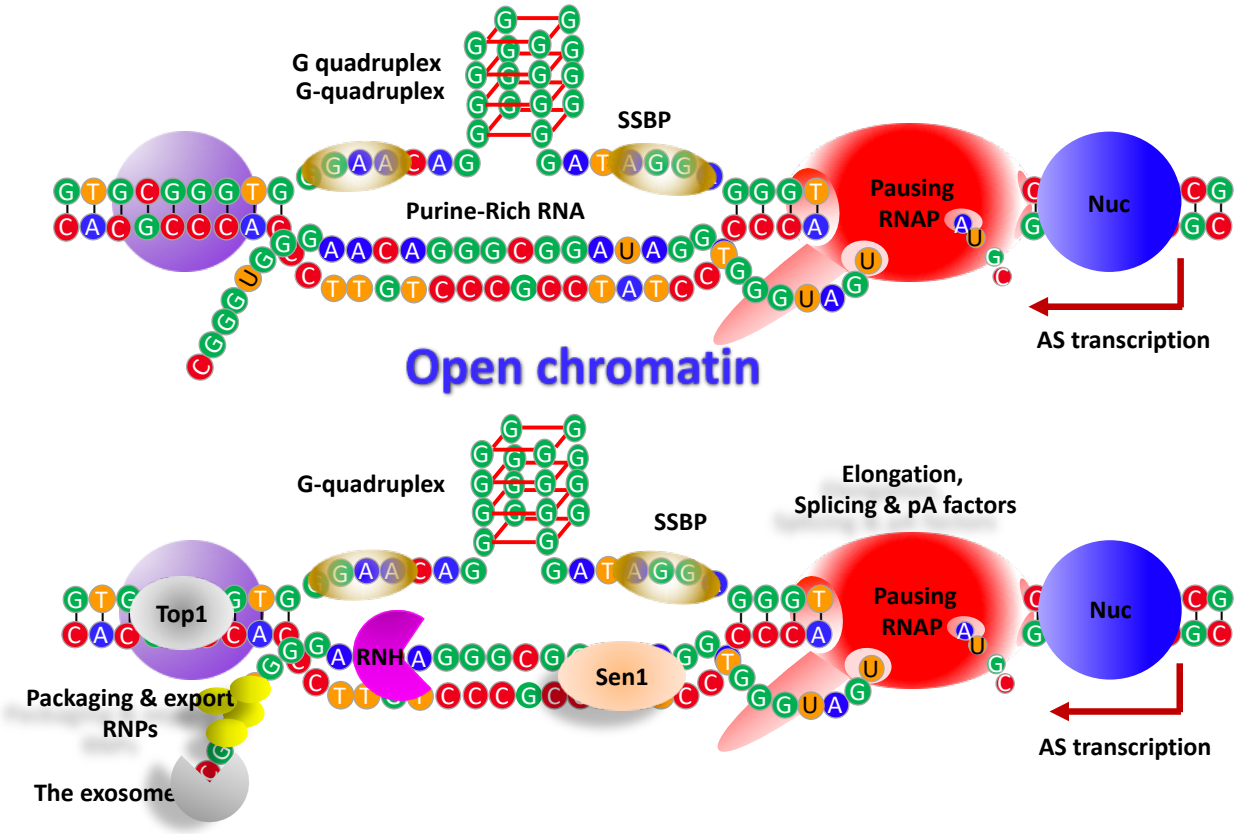



**Figure 1.1. Co-transcriptional R-loop formation and R-loop control factors.** Schematic representation for an R-loop forming co-transcriptionally, behind elongating RNA Pol and over an open-chromatin region with antisense transcription. Different factors such as RNase H, RNA-DNA helicase Sen1, topoisomerase Top1 and RNA processing machinery control R-loop formation. G quadruplexes and signle-stranded DNA-binding proteins are thought to stabilize R-loop formation. AS, antisense; Nuc, nucleosome; pA, polyadenylation; RNAP, RNA polymerase; RNH, RNase H; RNPs, ribonucleoproteins; SSBP, single stranded DNA binding protein.

While the mechanisms behind their formation aren't clear, multiple factors have been recognized to enhance formation of R-loops. Conversely, different factors monitor R-loops and control their levels to evade the deleterious consequences of their accumulation (Figure 1.1). R-loops range in size from few base pairs to few kilobases. An example for short R-loops (8-bp long) are those which form transiently to be used by the replication fork during lagging DNA strand synthesis. Another example for short R-loops are those forming inside elongating RNA polymerases during transcription. These are known to be around 17-bp long. Long R-loops are best represented by those forming over class switch recombination (CSR) regions in mammalian lymphocytes, and may extend to 2 kb in length, **(Crossley *et al.*, 2019; Garcia-Muse & Aguilera, 2019; Garcia-Pichardo *et al.*, 2017; Santos-Pereira & Aguilera, 2015)**.

## 1.2 R-loop discovery: a historical background

From a historical point of view, R-loops were shown to exist under artificial conditions before they were found to form *in vivo*. *In vivo* R-loops were first shown to form in *E. coli* and, later, in other biological systems.

### 1.2.1 Detection and characterization of artificial R-loops *in vitro*

R-loops were first discovered in a search for a tool to fractionate and separate DNA fragments of a specific sequence from total DNA. This has been achieved by annealing an RNA of complementary sequence to the target dsDNA in a way that maintains the integrity of the dsDNA and doesn't completely denature it **(Robberson *et al.*, 1972)**. The formed structure was studied using electron microscopy by White and Hogness who observed that the rRNA of *Drosophila* could hybridize to the template rDNA strand to form a three stranded structure similar to the D-loop (Expansion displacement loop) **(Robberson *et al.*, 1972)**. D-loops are three-stranded DNA structures that form transiently during mitochondrial DNA (mtDNA) replication**.** The main difference was that the duplex part in the newly-identified structure consisted of a DNA-RNA hybrid rather than DNA-DNA duplex in D-loop (Figure 1.2).

Davis lab further investigated the efficiency of RNA-DNA hybridization under different artificial conditions. Using scanning electron microscopy, they found that in 70% formamide buffer, RNA can efficiently hybridize to the template strand to form an RNA-DNA hybrid. This novel structure consisting of a displaced DNA strand and an RNA-DNA hybrid, was called as "R-loop" **(Thomas *et al.*, 1976)**. A year later, White and Hogness decided to use their

method to map R-loop formation over 18S and 28S rDNA of *Drosophila* by hybridizing the rRNA and rDNA in formamide buffer then detecting the formed R-loops under electron microscopy **(White & Hogness, 1977)**.

### 1.2.2 Early evidence for R-loop existence *in vivo*

R-loops were evidenced to exist *in vivo* for the first time in *E. coli* in 1994 by **Drolet *et al.* (1994)**. They have demonstrated that plasmids extracted from an *E. coli* strain lacking topoisomerase I (*topA*Δ**)** contained R-loops while plasmids extracted from a strain with the wild-type enzyme didn't. Hyper-negatively supercoiled plasmids extracted from *topA*Δ *E. coli* tended to run late on a chloroquine-containing agarose gel. Treating these plasmids with RNase H prior to gel separation, accelerated the run. Furthermore, treating the R-loop-containing hyper-negatively supercoiled plasmids with topoisomerase II caused the relaxation of these plasmids on gel, only, if treated with RNase H prior to topoisomerase II **(Drolet *et al.*, 1994)**.

The same group further investigated R-loops and their formation and found that overexpressing RNase H in *topA*Δ and growth-deficient *E. coli* can partially alleviate the growth defect caused by the loss of topoisomerase I in these bacterial strains. They concluded that R-loop formation is dependent on transcription and that the absence of topoisomerase I can induce their formation **(Drolet *et al.*, 1995)**. With the advent of R-loop-specific S9.6 antibody and the introduction of different R-loop detection methods, R-loops have been recently shown to form frequently and abundantly over the genome of many prokaryotic and eukaryotic organisms **(Crossley *et al.*, 2019; Garcia-Muse & Aguilera, 2019)**.

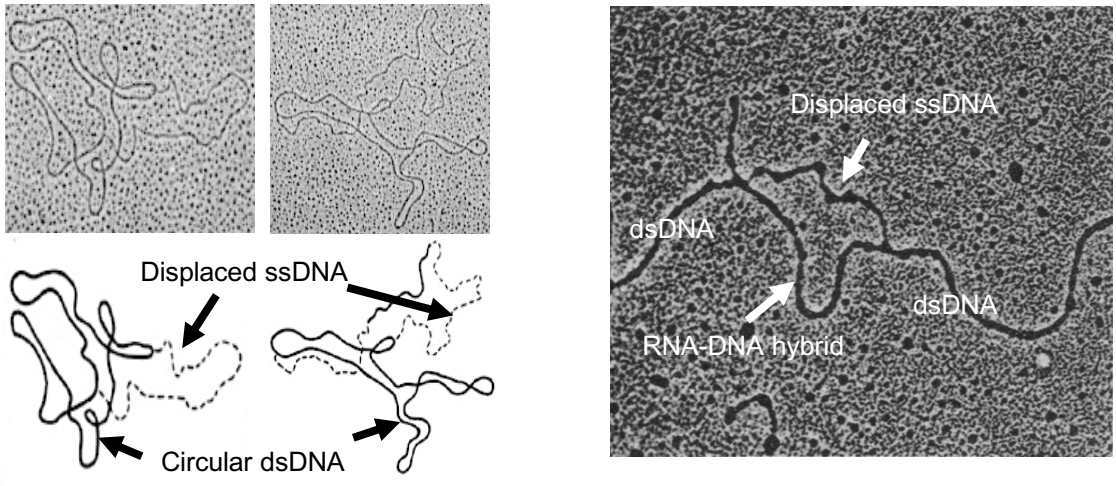



**Figure 1.2. Scanning electron micrographs for D-loops (left) and R-loops (right).** Adapted from **Robberson *et al*. (1972)** and **Thomas *et al*. (1976)**, respectively.

## 1.3 Models and mechanisms for R-loop formation

During transcription, RNA polymerases generate a copy of the genetic material in the form of an RNA strand complementary to the template DNA strand. This newly transcribed RNA (nascent RNA) is promptly released, processed, packaged and exported outside the nucleus. However, it may be retained in the nucleus, and invade a duplex DNA to competitively hybridize to the complementary DNA strand, taking the place of the non-complementary one of similar sequence. This process ends up with an RNA-DNA hybrid and an exposed ssDNA. The entire structure is termed an R-loop **(Belotserkovskii *et al.*, 2018)**.

### 1.3.1 Co-transcriptional R-loop formation (*in cis* R-loops)

The exact mechanism by which R-loops can form isn't clearly understood. According to some models, R-loops can form co-transcriptionally behind elongating RNA Polymerases through hybridizing to the template DNA strand to form an RNA-DNA hybrid. This class of R-loops are called *in cis* R-loops. As shown in Figure 1.3, there are three main models explaining how *in cis* R-loops may form; the threading back model, the extension model and the backtracking model **(Belotserkovskii *et al.*, 2018; Gowrishankar *et al.*, 2013)**.

#### *1.3.1.1 Strand extension or bubble extension model*

According to this model, short RNA-DNA hybrids that form inside the transcription bubble during transcription elongation may continuously extend behind the elongation complex to form extensive R-loops (Figure 1.3A). By another meaning, an RNA-DNA hybrid forming inside an RNA polymerase may not get separated behind the wake of transcription complex **(Belotserkovskii *et al.*, 2018; Gowrishankar *et al.*, 2013)**.

Contradictory to this model are the crystal structure as well as the cryo-electron microscopy studies for elongating RNA polymerases which show that each of the RNA and the DNA exits the elongating RNA polymerase through different channels **(Bernecky *et al.*, 2016; Farnung *et al.*, 2018; Hahn, 2004; X. Liu *et al.*, 2013; X. Liu *et al.*, 2018; Schier & Taatjes, 2020; Tafur *et al.*, 2016; Westover *et al.*, 2004)**. However, this model cannot be completely excluded.

#### *1.3.1.2 Threading back (reannealing model)*

According to this model, after the RNA exits the transcription bubble, it may thread back and reanneal with the template DNA to form an R-loop structure **(Belotserkovskii *et al.*, 2018; Gowrishankar *et al.*, 2013)**. Although this is the most accepted model for R-loops formation, it doesn't provide a deep insight about the core mechanisms for R-loop formation. It fails to explain whether this can happen directly once the RNA exits the transcription bubble, or whether it requires a long RNA molecule to initiate the process. If the later assumption were true, then how could the RNA get threaded, especially, in WT cells with fully functional RNA packaging and processing factors?

Besides, this model can't explain whether this process requires the DNA to be in a single stranded configuration or whether the RNA can thermodynamically compete with the non-template DNA strand and replace it. In different words, how the unwinding of DNA behind the RNA polymerase gets initiated is not clear. Interestingly, three situations could be envisaged for strand invasion by RNA (Figure 1.3B). Strand invasion may occur through inserting either the 5' end, the 3' end or the middle of the RNA into the duplex DNA **(Gowrishankar *et al.*, 2013)**. Intriguingly, additional mechanisms have been suggested for co-transcriptional R-loop formation through the threading back model.

#### *1.3.1.3 G clustering*

G clustering is an accessory model that complements the threading back model for R-loop formation. According to this model, G clustering is required for initiating and stabilizing R-loops formation **(Roy & Lieber, 2009; Roy *et al.*, 2008)**. The presence of G clusters at the initiation spot of R-loop formation over a class switch DNA construct, was found to be strictly required for R-loop initiation during *in vitro* transcription with T7 RNA polymerase. Without these clusters, R-loops formation was inefficient **(Roy & Lieber, 2009)**.

Meanwhile, the high G density was required for the elongation of R-loops, but not the initiation. This suggests that G clustering may be the trigger for RNA threading back and initiation of R-loop formation behind the elongating RNA polymerase. Once an R-loop is formed, it can elongate in a way dependent on high G density of the non-template DNA strand **(Roy & Lieber, 2009)**. According to the different strand invasion mechanisms, G-clustering triggers R-loops formation through strand invasion by the 5' end of RNA.

#### *1.3.1.4 DNA opening and strand exchange with PRC2*

A novel mechanism that has been recently postulated for R-loop formation is DNA opening and RNA-DNA strand exchange through Polycomb Repressive Complex 2 (PRC2) **(Alecki *et al.*, 2020)**. PRC1 and PRC2 are members of the Polycomb group (PcG) proteins which are known to recognize and bind RNA and DNA separately, and R-loops as well **(Skourti-Stathaki *et al.*, 2019)**. In *Drosophila*, they bind to specific DNA sequences called Polycomb Response Elements (PREs) to regulate gene expression and early embryonic development by writing the repressive histone mark H3K27me3 **(Alecki *et al.*, 2020)**.

Interestingly, both PRC1 and PRC2 were shown to open dsDNA and identify R-loops *in vitro*. Moreover, PRC2 induced R-loop formation from ssRNA and dsDNA under the same conditions **(Alecki *et al.*, 2020)**. This suggests a role for PcG proteins in R-loop formation *in vivo*. This hypothesis is supported by the confirmed enrichment of R-loops at 25% of PREs in *Drosophila* and the propensity of these R-loop-forming PREs to be more occupied by PcG proteins than non-R-loop-forming ones **(Alecki *et al.*, 2020)**.

Nevertheless, the strand exchange activity and R-loop formation by PRC2 necessitated the presence of a free 3' DNA end. This implies the requirement for DNA break which isn't possible *in vivo.* Perhaps, the contribution of PRC2 to R-loop formation *in vivo*, if assumed to be true, depends on the activity of other proteins that yet to be identified **(Alecki *et al.*, 2020)**.

#### *1.3.1.5 Backtracking model*

According to this model, the backtracking (backsliding) of the transcription elongation complex **(Dutta *et al.*, 2011; Nudler, 2012)** prior to release and termination can be another pathway for R-loop formation **(Gowrishankar *et al.*, 2013)**. Backtracking leads to unwinding the duplex DNA and reannealing of the RNA to the template DNA upstream the transcription bubble (Figure 1.3C). It also involves threading the ssRNA along the RNA polymerase and extending the short RNA-DNA hybrid inside the transcription bubble **(Gowrishankar *et al.*, 2013)**.

Evidence for backtracking as a cause of R-loop formation came from work on the two bacterial proteins, Rho and NusG. Rho is an RNA helicase that terminates transcription of non-simultaneously translated RNA while NusG is a factor required for this termination process

which is called Rho Dependent Transcription Termination (RDTT) **(Peters *et al.*, 2011; Rabhi *et al.*, 2010)**. Disturbing this RDTT pathway by introducing deficiency in either NusG or Rho leads to the accumulation of antisense transcripts and R-loops genome wide **(Leela *et al.*, 2013; Peters *et al.*, 2012; Raghunathan *et al.*, 2018)**.

Backtracking has been found to be a prominent feature of Pol II at non-simultaneously translated RNA in bacteria. Backtracking was found to induce replication-dependent DNA DSB damage mainly through accumulating R-loops over these regions. This represented the first evidence that backtracking induces R-loop formation **(Dutta *et al.*, 2011)**. Deleting either Rho or NusG causes lethality of bacterial cells due to severe accumulation of R-loops. Interestingly, introducing a mutation in RNA Pol II that made it resistant to backtracking rescued the lethal phenotype of Rho or NusG deletion **(Leela *et al.*, 2013)**.

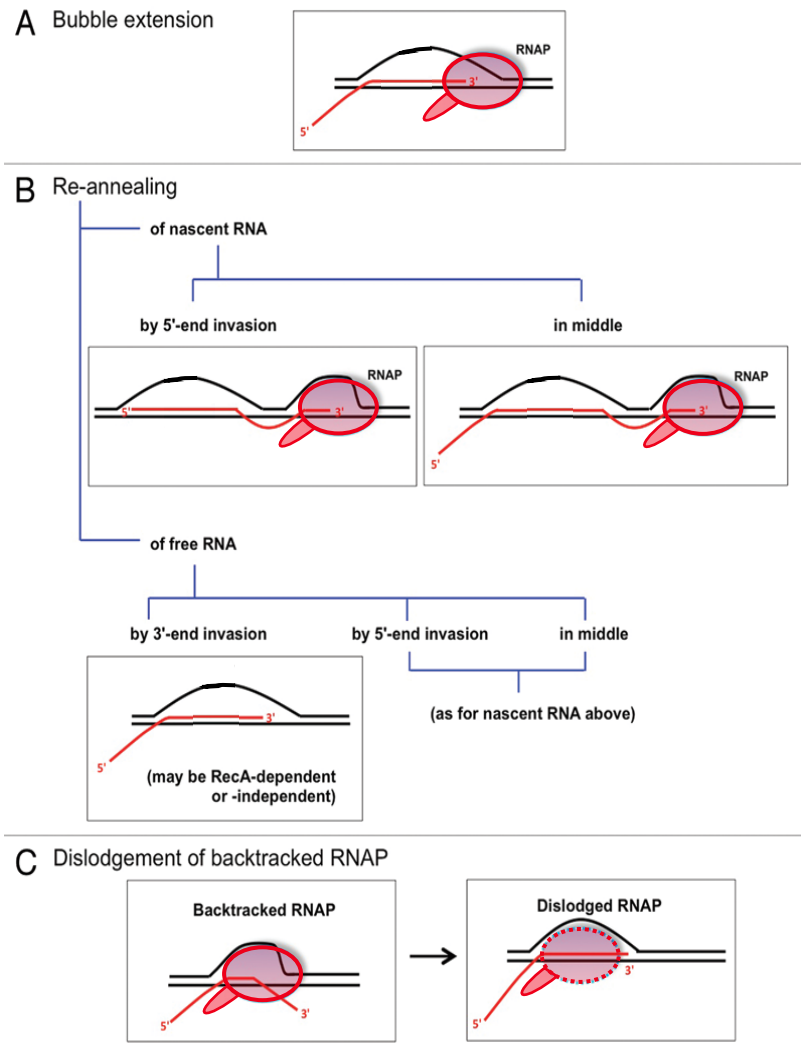



**Figure 1.3. Different models and strand invasion mechanisms suggested for R-loop formation.** Schematic representation for R-loop formation models and strand invasion by RNA end. **(A)** extension, **(B)** threading back and **(C)** backtracking models (see text for details). Adapted from **Gowrishankar *et al.* (2013)**.

This strongly supports backtracking as a model for R-loop formation, at least, in bacteria. However, the maximum length of R-loops attributed to backtracking was found to be 25 bp which makes this model limited to short R-loops **(Dutta *et al.*, 2011; Gowrishankar *et al.*, 2013; Nudler, 2012)**. This represents a drawback for this model as backtracking can't be an accepted model for the formation of long stretching R-loops over terminator regions in different biological systems.

In general, the *in cis* mechanism of R-loop formation is supported by the fact that most of the RNA biogenesis factors that have been implicated in preventing R-loop formation are known to function co-transcriptionally. However, it can't explain how R-loops can form and accumulate in mutants for RNA export and degradation enzymes which work post-transcriptionally **(Wahba *et al.*, 2011; Wahba *et al.*, 2013)**.

### 1.3.2 Post-transcriptional R-loop formation (*in trans* R-loops)

R-loops may also form post-transcriptionally (*in trans*) when RNA hybridizes to DNA at an ectopic region different from the region where the RNA was transcribed. Multiple factors and enzymes have been found to facilitate the formation of *trans* R-loops **(Belotserkovskii *et al.*, 2018)**.

#### *1.3.2.1 The prokaryotic CRISPR-Cas9 system*

The prokaryotic CRISPR-Cas9 system which is part of the bacterial immunity system provides an outstanding proof for the formation of R-loops *in trans*. The CRISPR-Cas9 system has been extensively used in genome editing and manipulation. This ribonucleoprotein complex uses a short guide RNA to recognize and hybridize to its target site by forming an RNA-DNA hybrid prior to cleaving the dsDNA **(F. Jiang & Doudna, 2017)**. The mechanism by which the short guide RNA can identify and initiate hybrid formation with the target DNA has been extensively investigated. It has been shown that the binding of Cas9 to the target DNA induces sharp twisting in the DNA and formation of a phosphate lock loop between Cas9 and the target DNA strand (Figure 1.4). This lock loop involves the interaction between the phosphate group of the phosphodiester bond immediately upstream the PAM sequence and specific residues in Cas9 (Figure 1.4). It's thought that this sharp twisting induces melting of DNA and invasion of RNA for R-loop formation **(F. Jiang & Doudna, 2017)**.

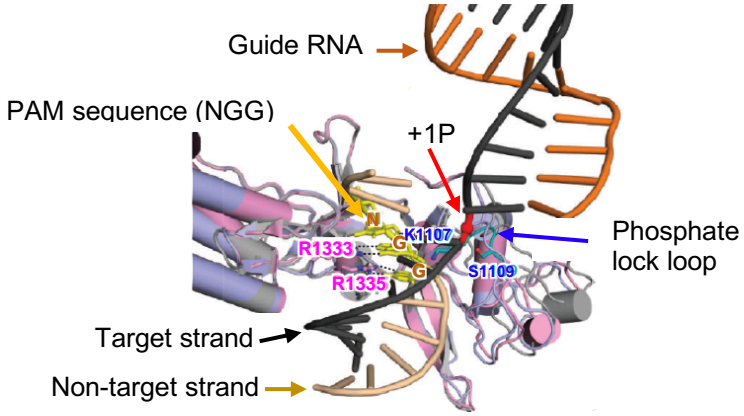



**Figure 1.4. Mechanism of R-loop formation by Cas9.** Close-up view showing induced sharp twisting in the DNA and formation of a phosphate lock loop (red sphere) between Cas9 and the target DNA strand (black strand). This lock loop involves the interaction between the phosphate group of the phosphodiester bond immediately upstream the non-target strand 5' NGG 3' PAM sequence (yellow colored sticks) and specific residues in Cas9. Adapted from **Jiang and Doudna (2017)**.

### 1.3.3 The homologous recombination and DNA damage repair machinery

#### *1.3.3.1 RecA*

Recombinase protein A (RecA), the bacterial orthologue of Eukaryotic Rad51 protein, is an essential component of the homologous recombination and DNA damage repair in bacteria. It has been reported to induce RNA-DNA exchange and RNA-DNA hybrid formation *in vitro* **(Kasahara *et al.*, 2000; Zaitsev & Kowalczykowski, 2000)**. This exchange process has been found to, crucially, require the presence of a mismatch in the duplex DNA (heteroduplex DNA) to trigger the action of RecA and initiate the displacement of the DNA by the single-stranded RNA *in vitro* **(Kasahara *et al.*, 2000)**.

RecA has been found to initiate RNA-DNA hybrid formation through a recombination process different from the conventional mechanism. RecA is traditionally known to facilitate the invasion of ssDNA into dsDNA. However, it was shown to induce an inverse strand exchange between RNA and DNA by first forming a nucleoprotein filament with dsDNA then replacing a DNA strand with the ssRNA *in vitro* **(Zaitsev & Kowalczykowski, 2000)**. It has been postulated that R-loops formation through RecA can present an alternative mechanism for DNA replication that depends on DNA-RNA recombination **(Kasahara *et al.*, 2000; Zaitsev & Kowalczykowski, 2000)**. However, this intriguing hypothesis has never been tested *in vivo*.

#### *1.3.3.2 Rad51*

Rad51 protein is known to facilitate the DNA-DNA exchange during chromosomal homologous recombination. Intriguingly, it has been shown to promote the formation of RNA-DNA hybrids and R-loops in budding yeast *Saccharomyces cerevisiae* **(Wahba *et al.*, 2013)**. Rad51 has been suggested to induce the formation of DNA-RNA hybrid *in trans* by facilitating the replacement of DNA by an RNA strand that was transcribed at another chromosomal locus homologous to the hybrid-forming locus **(Wahba *et al.*, 2013)**. However, and as usual, there is a missing piece for the jigsaw. Although work on Rad51 enlightened how R-loops can form *in trans* and *in vivo*, it's not known how Rad51 can direct the exchange process between DNA and RNA for promoting R-loop formation. Does Rad51 work through the canonical direct reaction by first binding to ssRNA or through the inverse direction described for RecA?

A striking example for an RNA inducing *trans* R-loop formation is the long non-coding RNA *APOLO* (*Auxin-Regulated Promoter Loop*). Most recently, it has been found to identify and target distant genomic regions through short homology search and base-pairing to form *trans* R-loops at these loci. These *trans* R-loops modulate the expression and function of auxin-responsive transgenes in *Arabidopsis* during root development **(Ariel *et al.*, 2020)**.

In addition to inducing post-transcriptional (*trans*) R-loop formation, the homologous recombination machinery may represent another pathway for the formation of co-transcriptional R-loops as well **(Costantino & Koshland, 2015; Wahba & Koshland, 2013)**.

One putative mechanism for R-loop formation, yet has never been tested, is that the presence of R-loop at a specific locus can render the non-template DNA strand free and accessible for RNAP to initiate antisense transcription and R-loop formation over the reverse strand. The final outcome would be a double R-loop, i.e, an R-loop might drive the formation of another R-loop in the opposite direction.

## 1.4 Biophysical and biochemical properties of R-loops and R-loop-forming regions

Beside their biological and functional relevance, RNA-DNA hybrids and R-loops are also regarded for their potential targeting and therapeutic applications. Therefore, understanding their biophysical and biochemical properties including topography (multi-stranded nature), thermodynamic stability and conformation of the constituting RNA-DNA hybrid, along with other properties, is of crucial importance.

These properties dictate the behavior of R-loops *in vitro* and *in vivo*. By their turn, R-loop formation affects the features and the local genomic landscape of R-loop-forming regions *in vivo*. For example, R-loop formation induce changes in chromatin features such as histone marks, histone occupancy, DNA replication, RNA transcription and Pol II kinetics. R-loop formation may also affect the sensitivity of the forming sequences to DNases, ribonucleases and different nucleases. In fact, R-loops are known to be associated with open chromatin, DNase-I hypersensitive regions and CpG methylation-free promoters **(Sanz *et al.*, 2016)**. In simpler words, R-loops form as a consequence for specific inherent features of the R-loop forming sequences and, mutually, R-loops impart specific changes and affect the response and activities of these regions to different cues.

### 1.4.1 Topography of R-loops

Nucleic acids exist in many forms, either in single-, double- or multi-stranded forms. They can base pair and interact through Watson and Crick hydrogen bonding to form homoduplexes or heteroduplexes (hybrids). Moreover, nucleic acids may interact through Watson-Crick or Hoogsteen hydrogen bonding to form triplexes such as H-DNA or quartets (quadruplexes) such as G-quadruplexes (Figure 1.5A) **(Belotserkovskii, Mirkin*, et al.*, 2013; Tateishi-Karimata & Sugimoto, 2014; G. Wang & Vasquez, 2014, 2017)**.

Intramolecular triplex DNA structures, such as H-DNA triple helix, have been shown to exist at DNA sequences rich in homopurine-homopyrimidine repeats **(Holder *et al.*, 2015; Kohwi & Kohwi-Shigematsu, 1988; Mirkin, 2008; Mirkin *et al.*, 1987)**. H-DNA triplex structure forms when a single DNA strand of inverted mirror-repeat sequence separates from its complementary strand and bends to bind with duplex DNA of the same repeat sequence on the same DNA molecule. The generated free DNA strand contributes to the sensitivity of the

H-DNA structure to S1 nuclease treatment **(Kohwi & Kohwi-Shigematsu, 1988; Mirkin *et al.*, 1987)**.

The H-DNA structure was initially shown to form under strong pH conditions where the third strand in the triplex was a pyrimidine strand forming Hoogsten hydrogen bonds with the purine strand in the duplex DNA, referred to as T*A-T and C*G-C **(Mirkin *et al.*, 1987)** as shown in Figure 1.5A. Later, another isoform of H-DNA was shown to form at weak acidic to normal pH and to require Magnesium ions. The third strand in this isoform was a purine strand making reverse Hoogsteen bonds with the purine strand in the DNA duplex, referred to as G*G-C and A*A-T **(Kohwi & Kohwi-Shigematsu, 1988)**. It's worthy mentioning that negatively supercoiled DNA facilitates the formation of H-DNA triplex structure **(Kohwi & Kohwi-Shigematsu, 1988; Mirkin *et al.*, 1987)**.

Interestingly, triplexes have been shown to form *in vivo* as well. For instance, Human *c-Myc* promoter has been shown to have a triplex-forming sequence that causes partial transcriptional arrest when T7 RNA Polymerase transcribes through the nuclease sensitive element of this promoter. This was supported by the finding that the intensity of the arrest increases with the increase of negative super-helicity **(Belotserkovskii *et al.*, 2007)**.

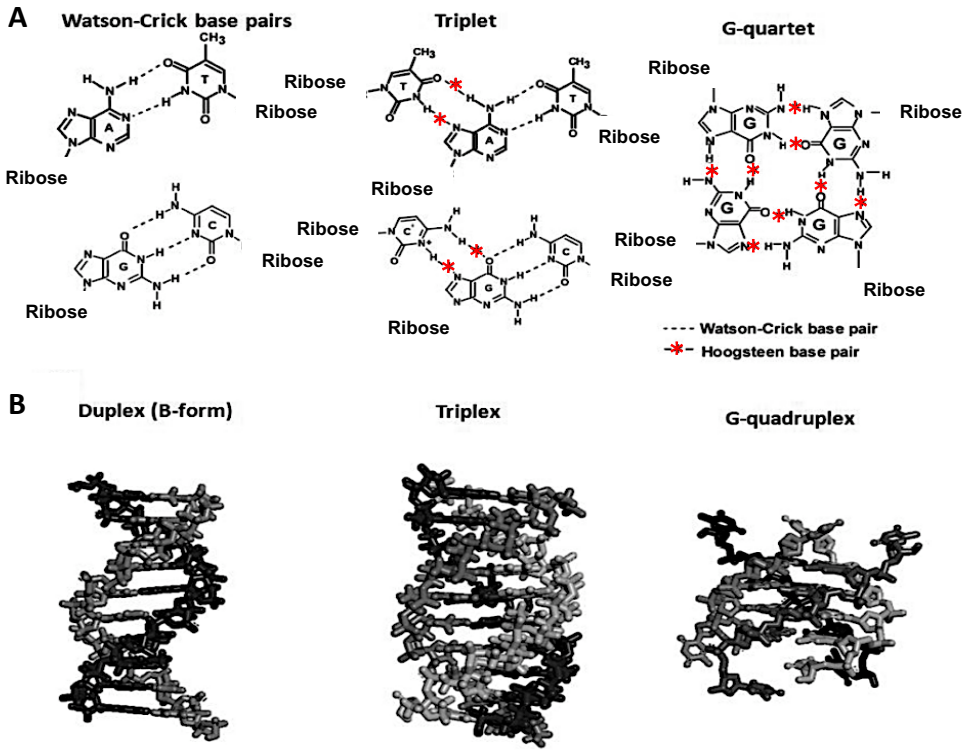



**Figure 1.5. Different types of bonds found in different nucleic acids at acidic conditions.** Adapted from **Tateishi-Karimata and Sugimoto (2014)**.

Comparable to H-DNA, can ssDNA interact with RNA-DNA hybrid in an R-loop and form a triplex? In fact, reports about triplex R-loops are very scarce and not supported by strong evidences. A structure called "collapsed R-loop" was suggested to form over a murine class switch region by *in vitro* transcription. It has been suggested that the RNA can reside in the major groove of the duplex DNA to form a triplex structure **(Reaban *et al.*, 1994)**. Triplex R-loop formation has been suggested as a model for blocking the poly-purine tract (PPT) and inhibiting reverse transcription in HIV1 viral RNA **(Shaw & Arya, 2008; Volkmann *et al.*, 1995)**. Interestingly, triplex R-loop formation at rRNA promoter has also been hypothesized as a mechanism mediating human rRNA gene silencing **(Schmitz *et al.*, 2010)**. However these hypotheses haven't been validated enough as the nature of these triplex structures hasn't been verified **(Belotserkovskii, Mirkin*, et al.*, 2013)**.

Based on electrophoretic mobility and circular dichroism (CD) spectra analysis, it has been inferred that a short duplex DNA, connected by a hair pin at one end, can form a triple helix with a short ssRNA *in vitro* **(Roberts & Crothers, 1992)**. Different hair-pin-connected duplex and heteroduplex oligonucleotides [either Purine DNA-Pyrimidine DNA (DD), Purine RNA-Pyrimidine RNA (RR), Purine DNA-pyrimidine RNA (DR) or Purine RNA - Pyrimidine DNA (RD)] were annealed with either Pyrimidine ssDNA (D) or Pyrimidine ssRNA (R) to check the possibility of triplex formation by these combinations. Six of these combinations were suggested to form a triple helix, among them were the R + (DD) and D + (DR), while. Neither could D + (RD) nor D + (RR) form a triplex **(Roberts & Crothers, 1992)**. Despite these findings, no high-resolution detection or enzymatic validation methods were used to support these conclusions about triplex formation by short nucleotides.

Electron microscopic observations for *in vitro* R-loops have provided a more solid idea about R-loops topography as three stranded structures rather than triplexes **(Backert, 2002; Glover & Hogness, 1977; Thomas *et al.*, 1976; White & Hogness, 1977)**. Since their discovery and characterization, there has been a strong consensus that R-loops consist of an RNA-DNA heteroduplex and a free ssDNA. This is also much supported by the fact that the most popular and widely used antibody for R-loop detection and immunoprecipitation, the S9.6 antibody, was generated against an RNA-DNA hybrid not a full R-loop construct **(Boguslawski *et al.*, 1986)**.

### 1.4.2 Thermodynamic Stability of RNA-DNA hybrids and R-loops

The physical and chemical properties, including thermodynamic stability, of any nucleic acid is reliant on its base content, conformation, ionic strength and hydration **(Kankia & Marky, 1999; Saenger, 1984; Shaw & Arya, 2008)**. As heterogenous three-stranded nucleic acid structures, R-loop stability is reliant, in addition to the aforementioned factors, on the difference in thermodynamic stability between the derivative RNA-DNA hybrid and the original native dsDNA. For stabilizing an R-loop, the associations between the RNA and the complementary template DNA strand should be stronger than those between both strands of the duplex DNA. This should maintain the formation of a stable RNA-DNA hybrid which guarantee the stability of the whole R-loop structure. Destabilizing the RNA-DNA hybrid would compromise the R-loop. The stability of the RNA-DNA heteroduplex itself is reliant, in the first instance, on the nucleotide sequence of the RNA and DNA strands assuming that the other variables above as constants while comparing different structures. Understanding the thermodynamic stability of R-loops would be important for understanding the mechanisms for their formation and resolution. A lot of efforts have been done to understand the correlation between base composition and thermodynamic stability of RNA-DNA hybrid.

Earliest trials for routine annealing of ribosomal RNA-DNA were first done in a 2X or 6X SSC media (1X SSC is 0.15 M NaCl, 0.015 M sodium acetate, pH 7.2) at 67 °C. Temperature above 80 °C was found to be required to increase the rate and efficiency of hybridizing ribosomal eukaryotic RNA to DNA in such buffer. Yet, this high temperature had deleterious impacts on the stability and integrity of the RNA which necessitated the use of another buffer to eliminate the need for high temperature. Surprisingly, using formamide increased the rate and efficiency of hybridization, eliminating the need for high temperature and maintaining the integrity of RNA **(Birnstiel *et al.*, 1972)**.

**Birnstiel *et al.* (1972)** found that in 0.1X SSC buffer, the melting temperature of *Xenopus* ribosomal RNA-DNA hybrids was lower than or equal to that of the corresponding DNA. This was consistent with the general belief at this time that the RNA-DNA hybrid is less stable than the native DNA. On the contrary, they found the melting temperature of the same RNA-DNA hybrid to be higher than that of the corresponding native DNA of the same GC content in 6X SSC-50% formamide buffer. This observation provided a strong evidence that formamide increases the efficiency of hybridization and the stability of RNA-DNA hybrid as well

**(Birnstiel *et al.*, 1972)**. Based on multiple findings, they concluded that different features of the RNA strand, such as its length, GC content, base-pairing efficiency and secondary structures, define the stability of the RNA-DNA hybrid. For instance, high-molecular-weight RNA-DNA hybrids are more stable than short RNA-DNA hybrids.

Since then, formamide buffer has been used for efficient R-loop preparation. In fact it has been reported that more than 140 studies used this method for artificial R-loop preparation **(Maizels *et al.*, 2014)**. One of the early studies, directed to investigate R-loop formation in formamide, has shown that the highest rate of R-loop formation occurs at the denaturing temperature of the R-loop-forming dsDNA. The same study found that RNA-DNA hybrid has more thermodynamic stability near this temperature compared to dsDNA in 70% formamide buffer **(Thomas *et al.*, 1976)**. Notably, this study used a viable system to facilitate R-loop formation; 2.6 kb of yeast rDNA abended to λ phage DNA was annealed to 26S yeast rRNA in 70% vol/vol formamide. The RNA-DNA hybrid stayed intact even after removing formamide, but the stability decreased. By increasing temperature, the dsDNA could be restored as a result for branch migration process. However, this process was slow for long RNA-DNA molecules that the degradation of RNA part was required for restoring the dsDNA **(Thomas *et al.*, 1976)**. This study concluded that R-loop formation process requires a very complex reaction whose rate is a function of reaction temperature, RNA concentration and ionic strength, and, theoretically, other variables such as length, secondary structures, and base composition of the RNA strand **(Thomas *et al.*, 1976)**.

Although the two studies conducted by **Birnstiel *et al.* (1972)** and **Thomas *et al.* (1976)** tended to change the belief about RNA-DNA hybrid and R-loop stability, respectively, their findings can't be accepted as a rule for *in vivo* conditions. First, their artificial system used formamide which may have differential effects on the stability of different nucleic acids. Formamide is a non-physiological denaturing agent which significantly destabilizes dsDNA and decreases the re-association of DNA strands if present in such high concentration **(Blake & Delcourt, 1996; Fuchs *et al.*, 2010; Sadhu *et al.*, 1984)**. Second, they mostly used long rRNA-rDNA fragments which may have specific features, sequence repeats and GC content that would support R-loop formation in general. In fact, **Thomas *et al.* (1976)** protocol required incubating the samples for 20 hours to ensure maximal rate of R-loop formation, which means the process isn't as easy as seems to be. Since they haven't tried different nucleic acid fragments, they didn't reveal the correlation between base content and stability.

Multiple studies further investigated the correlation between base content and hybrid stability. It has been suggested that purine RNA-pyrimidine DNA hybrids and pyrimidine RNA-purine DNA hybrids belong to different structural categories and have different stabilities. Comparing relative stability of different duplex homopurine (pu)-homopyrimidine (py) oligonucleotides, their relative stabilities were found to be in the order of R(pu)-R(py) > R(pu)-D(py) > D(pu)-D(py) > D(pu)-R(py) **(Hung *et al.*, 1994; Roberts & Crothers, 1992)**. This implies that nucleotide base content of the RNA defines the stability of the RNA-DNA hybrid. The stability and strong binding of R(pu)-R(py) and R(pu)-D(py) oligonucleotides explains why neither D(py) + R(pu)-D(py) nor D(py) + R(pu)-R(py) oligonucleotide combinations could form triplex structures (**Roberts & Crothers, 1992)** (Section 1.4.1) **.** Using RR, DD, RD and DR duplexes with the sequence CCATCGCTACC representing the pyrimidine strand and GGTAGCGATGG representing the purine strand, and with U instead of T for RNA, the same RR > RD> DD> DR order of stability was confirmed **(Kankia & Marky, 1999)**.

This correlation has been also checked for short sequences. The above trend was shown to change in a study using short pentamer duplexes of mixed purine-pyrimidine sequences (GT-CA) corresponding to T stem of Phenylalanine tRNA. RR duplex was the most stable, while the stability of both RD and DR was less than DD, with R(GU)-D(CA) more stable than D(GT)-R(CA) **(Hall & McLaughlin, 1991).**

Two groups of different hybrids were used as models to study relative stabilities of different RNA-DNA hybrids. The first group consisted of four possible homopurine-homopyrimidine (DD, DR, RD and RR) hybrids of the sequence (GAAG)3-(CTTC/CUUC)3, while the second consisted of mixed purine-pyrimidine hybrids (DD, DR, RD and RR) of the sequence (GTTG/GUUG)3-(CAAC)3. Consistent with **Roberts and Crothers (1992)** and **Hung *et al.* (1994)**, the stability of the first group was found to be in the order of RR > RD > DD > DR while, close to **Hall and McLaughlin (1991),** the stability of the second group was found to be in the order of RR > DD > D(GT)-R(CA) > R(GU)-D(CA) **(Ratmeyer *et al.*, 1994)**.

Indeed, these studies showed a common trend, and a conclusion could be drawn for the effect of sequence composition and RNA-DNA hybrid stability. As a rule, the prevalence of purine stretches in the RNA strand makes the RNA-DNA hybrid thermodynamically more stable than the DNA duplex. On the other hand, the presence of pyrimidines in RNA reduces

the hybrid stability to be less than that the DNA duplex. Nevertheless, all these studies depended on artificial buffers which may impact the relevance of their observations.

A comprehensive study was conducted to investigate the correlation between deoxy-pyrimidine (dPy) content, A-T/U tracts and length of different duplexes (RR, RD, DR, DD), and their thermodynamic stability in near physiological conditions, based on CD spectra and electrophoretic mobility **(Lesnik & Freier, 1995)**. 14 different duplex- and hybrid-forming sequences of variable lengths (8 – 21 bp) were used. Each sequence had four homologues (RNA-RNA, RNA-DNA, DNA-RNA and DNA-DNA). These sequences were divided into three groups based on the dPy (0 – 100%) distribution and A-T/U (25 – 100%) base content in the sequence of the DNA-RNA hybrid. The first group had 3 sequences of 0 : 100% dPy distribution (homopurine-homopyrimidine) and 33 - 44% A-T/U content, the second had four sequences of ~25 : 75% dPy distribution and 30 – 73% A-T/U content while the third had 5 sequences of ~50 : 50% dPy distribution and 25 – 80% A-T/U content. A fourth group was assigned separately based on poly-A tracts, with 2 sequences of 33 and 62% n(dA).

Regarding thermodynamic stability, this study found that; all dsRNAs were more stable than dsDNAs of the same sequence except the dsDNA of 62% n(dA); the increase of A-T/U content decreased the stability of all duplexes; the increase of dPy percentage in the DNA (increase of Pu in RNA) of the hybrid increased the stability of the hybrid compared to the dsDNA; and 50% dPy content is the point of "break-even" where the thermodynamic stability is even for RD, DR and DD. Considering hybrid geometry, they concluded that; RD hybrids (high dPy) are structurally different from DR (low dPy); RD and DR hybrids of 50% dPy have a structural geometry in between dsDNA and dsRNA; and the increase of dPy content increased the RNA-like conformation of the hybrid. Finally, regarding electrophoretic mobility on 20% PAGE, all dsRNA duplexes ran slower than dsDNA duplexes of the same sequence, and hybrids with high dPy content migrated closer to dsRNA while hybrids with low dPy content migrated closer to dsDNA **(Lesnik & Freier, 1995)**.

Interestingly, these correlations didn't fit for the poly-A duplexes and hybrids. The 33 and 66% n(A-T) duplex DNAs showed increased stability compared to the rest of mixed DNA sequences, and the stability of the 66% n(A-T) duplex DNA was substantially higher than the stability of the duplex RNA and hybrids of similar sequence. All the 66% n(A-T/U) duplexes and hybrids had CD spectra different from those of the other studied sequences and, finally, 33

and 66% n(rA-dT) hybrids migrated extraordinarily fast on PAGE compared to other sequences **(Lesnik & Freier, 1995)**.

Unfortunately, information about the stability of RNA-DNA hybrids in R-loop context isn't available. It would be important to study the stability of RNA-DNA hybrids in such context since the RNA-DNA hybrid is usually flanked by long dsDNA. The flanking dsDNA has different physio-chemical properties, conformation and super-helicity which exert stress on the R-loop to extrude the RNA, reanneal with the free DNA through branch migration and restore the default configuration of DNA. This reannealing is critical for maintaining double-stranded nature and integrity of DNA which is required for chromatin assembly and preventing DNA damage. It drives curiosity to think how R-loops would form and behave in chromatin environment (in the presence of nucleosomes).

The free DNA strand can be another important factor as well since it can, by somehow, interact with the RNA-DNA hybrid, affecting its stability and the overall stability of the R-loop. I'm not aware of any studies that have been conducted to address this point. All the studies in this section tried to provide an idea about relative stabilities of different hybrids in completely separate reactions and using short duplexes. Overall, these findings would be very helpful to predict the chance of R-loop formation and how stable it would be at a specific genomic region *in vivo*. Computational analysis for *in situ* mapped R-loops would help fill this gap and complement the picture.

### 1.4.3 Conformation/structural geometry of RNA-DNA hybrids

Nucleic acids are known to exhibit different spatial conformations. dsDNA has three structural geometries, the right-handed A-form with C3'-endo, the right-handed B-form with C2'- endo sugar pucker, and the left-handed Z form. The RNA dominantly exists in A-form with transition to other forms under specific conditions. Different parameters including nucleotide sequence, hydration and buffer conditions influence conformation and mediate transition among different geometries **(Belmont *et al.*, 2001; Choi & Majima, 2011; Murthy *et al.*, 1999; Saenger, 1984; Schneider *et al.*, 2004; Svozil *et al.*, 2008; Szabat & Kierzek, 2017)**. The conformation of different RNA-DNA duplexes has been extensively studied using different techniques which led to differences in observations and discrepancies in the findings of different studies regarding RNA-DNA hybrid conformation **(Shaw & Arya, 2008)**.

Different conformations of sugar puckers identified by multiple studies are shown in Figure 1.6.

The earliest study used X-ray fiber diffraction method to study the conformation of bacteriophage f1 DNA-RNA hybrid **(Milman *et al.*, 1967)**. This study found that changing hydration level didn't induce significant changes in the conformation of the hybrid. It also suggested that the RNA-DNA hybrid has a conformation more similar to A-form of DNA but different from the control viral dsRNA. Later, hydrated RNA-DNA hybrids were suggested to exhibit a conformation more similar to A-form of double-stranded RNA than B-form of double-stranded DNA **(O'Brien & MacEwan, 1970)**. This conclusion, based on X-ray filament diffraction pattern for poly rI-poly dC hybrid in solution, was supported by circular dichroism (CD) **(Gray & Ratliff, 1975)** and nuclear magnetic resonance (NMR) **(Pardi *et al.*, 1981)**. Investigations using X-ray crystallography for RNA-DNA hybrids further enhanced this belief **(Egli *et al.*, 1992; A. H. Wang *et al.*, 1982)**.

On the other hand, studies using X-ray diffraction **(Arnott *et al.*, 1986; Zimmerman & Pheiffer, 1981)**, NMR **(Fujiwara & Shindo, 1985; Shindo & Matsumoto, 1984)** and Raman spectroscopy **(Benevides & Thomas, 1988)** found that RNA-DNA hybrids, in solution, have bimorphic conformation with A-form RNA strand of C3'-endo ribose sugar (also known as north or C2'-exo) and B-form DNA strand of C2'-endo deoxyribose sugar (equivalent to south or C3'-exo). Using high-resolution 2D NMR confirmed that in solution, the RNA strand of RNA-DNA exists in A-form with C3'-endo ribose sugar (N-type) while the DNA strand exists in B-form (S-type), but the deoxyribose sugar doesn't have the typical S-type C2'-endo **(Chou *et al.*, 1989)**.

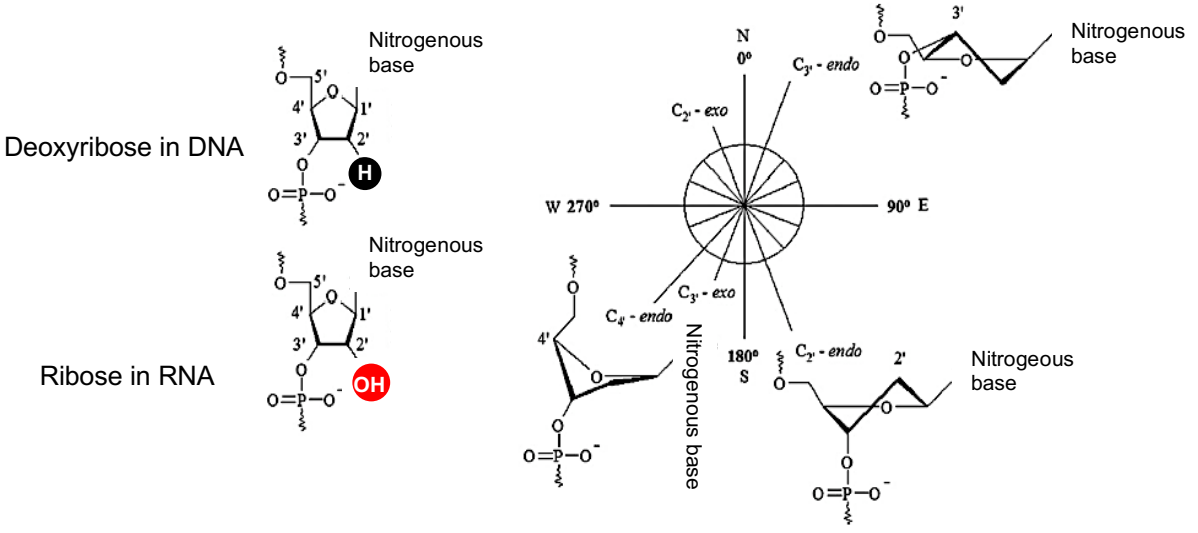



**Figure 1.6. Conformations of sugar puckers in RNA-DNA hybrids.** Adapted from **Shaw and Arya (2008)**.

Later, the DNA strand of an RNA-DNA hybrid was identified to neither exist in North A-form nor in south B-form, but in a spectacular O4'-endo conformation. This was confirmed by two-dimensional nuclear overhauser effect (2D-NOE) spectra **(Fedoroff *et al.*, 1993; Salazar *et al.*, 1993)**. Interestingly, these findings provided explanations why RNase H enzymes can bind dsRNA and RNA-DNA hybrid but not dsDNA, and can cleave RNA in hybrid only, but not in pure dsRNA. Variations in minor groove, major groove, twisting angle and hydration among dsDNA, dsRNA and hybrid are thought to be the reason for the differences in binding (recognition) and enzymatic activity (processivity) of RNase H on these different substrates **(Noy *et al.*, 2005)**.

Differences in electrophoretic mobility and circular dichroism spectra among different oligomeric duplexes of either RNA-RNA, RNA-DNA and DNA-DNA nature provided a very strong evidence that RNA-DNA hybrids exhibit structural geometries and have characteristics intermediate between pure RNA and DNA duplexes **(Bhattacharyya *et al.*, 1990; Hung *et al.*, 1994; Lesnik & Freier, 1995; Ratmeyer *et al.*, 1994; Roberts & Crothers, 1992)**.

Although the conformation of RNA-DNA hybrids in R-loop context hasen't been resolved or studied, in most cases RNase H can cleave the RNA strand of RNA-DNA hybrid in R-loops (RNase H-sensitive R-loops). This suggests that the conformation of hybrids in three stranded R-loop structure doesn't change compared to double stranded hybrids, or there might be very neglegible changes.

### 1.4.4 Electrophoretic mobility of RNA-DNA hybrids and R-loops

In neutral to weak alkaline medium, nucleic acids have a negative charge on the phosphate groups of the ribose sugar backbone. Since the net negative charge of nucleic acids is correlated with the total number and correspondingly the total mass of nucleotides, the total charge per mass ratio is constant for nucleic acids of similar structure. That's why gel separation of structurally-similar nucleic acids is based mainly on their size whereas their electrophoretic mobilities are affected by their molecular weights, i.e., low molecular weight molecules run faster than high molecular weight ones **(Ylmaz *et al.*, 2012)**.

Because it has been thought to have an A-form conformation similar to dsRNA, the RNA-DNA hybrid was expected to run parallel to dsRNA on gel. Unexpectedly, **Bhattacharyya *et al.* (1990)** found that RNA-DNA hybrids run at a distance between dsDNA

and dsRNA in 15% polyacrylamide gel, with dsRNA having the slowest mobility (highest profile). This was a simple and convincing evidence that the hybrid structure isn't similar to the structure of its pure components, but rather a mix between their structures.

**Roberts and Crothers (1992)** found that the sequence of the RNA not only defines the stability of the RNA-DNA hybrid but also affects its electrophoretic mobility. Using homopurine-homopyrimidine hair-pin oligomeric duplexes, they found that in 20% polyacrylamide gel, RR has the slowest mobility, DD has the fastest mobility, while the RD and DR run in between, i.e., RR runs at the highest level while DD runs at the lowest level. There was a striking difference between the mobility of RD and DR. RD ran slower (higher) and closer to RR while DR ran faster (lower) and closer to DD. This suggested that RD has a structure more similar to RR while DR has a structure closer to DD **(Roberts & Crothers, 1992)**. Interestingly all the triplex-forming combinations, including potential R-loop-forming ones, ran comparatively higher than all the duplexes and showed differences in their mobility imparted by differences in their sequences.

The same mobility speed (RR < RD < DR < DD) for homopurine-homopyrimidine oligomeric duplexes has been confirmed by an independent study of **Ratmeyer *et al.* (1994)**. Surprisingly, they found mixed-purine-pyrimidine duplexes to exhibit the same behavior (RR < R(GU)-D(CA) < D(GT)-R(CA) < DD) which confirmed that RD and DR are structurally different from each other and from DD and RR **(Ratmeyer *et al.*, 1994)**.

Finally, Circular dichroism spectra of homopurine-homopyrimidine duplexes supported the fact that RD has more structural similarity to dsRNA while DR has more similarity to dsDNA. Overall, this suggests that each of RD and DR belongs to different structural categories and neither has a pure A-from of dsRNA nor a pure B-from of dsDNA, but rather an intermediate structure **(Hung *et al.*, 1994; Ratmeyer *et al.*, 1994; Roberts & Crothers, 1992)**.

Knowing that the RNA strand in the hybrid has a conserved A-form for both purines and pyrimidines, the intriguing conclusion is that the conformation of the hybrid depends on the nucleotide content of the DNA. More precisely, pyrimidines in the DNA of RD tend to have a conformation closer to A-form which gives the hybrid a dsRNA-like conformation, while purines in DNA of DR exhibit a geometry more similar to B-form to give the hybrid a more dsDNA-like structural geometry.

## 1.5 Factors enhancing and stabilizing R-loop formation

Different factors have been identified to promote R-loop formation. Of these, the most determining factor is transcription rate. Besides, R-loop formation is enhanced by other factors that favor the associations between the RNA and complementary DNA over those between the two DNA strands. These include inherent features of the DNA sequence and the corresponding RNA, secondary structures formed in the non-template DNA strand, negative supercoiling and the presence of a nick in the non-template DNA strand. Moreover ssDNA-binding proteins stabilize R-loops through binding to the ssDNA and blocking its rehybridization to the template DNA strand **(Aguilera & Garcia-Muse, 2012; Al-Hadid & Yang, 2016; Allison & Wang, 2019; Chedin, 2016; Costantino & Koshland, 2015; Stephan Hamperl & Cimprich, 2014; Santos-Pereira & Aguilera, 2015; Skourti-Stathaki & Proudfoot, 2014)**.

### 1.5.1 Enhancing factors: intrinsic sequence features

R-loop formation is an inherent programmed feature of the genome of many organisms. The sequence of a specific genomic region is an important determinant for the R-loop forming potential of that region. This was strongly evidenced by an early study which found that a 140-bp RNA-DNA hybrid formed over an *in vitro*-transcribed murine IgA switch region. Importantly, this RNA-DNA hybrid formed regardless of the topological state (supercoiling) of the plasmid vector. Once formed, this hybrid was so thermodynamically stable that it tolerated temperatures above 95 °C and positive supercoiling stress of the vector **(Reaban *et al.*, 1994)**. Another study found that R-loops form *in vitro* over murine class switch regions only when transcription goes in the sense physiological but not in the antisense direction **(Daniels & Lieber, 1995)**. This strongly suggests that there is a direct correlation between R-loop formation and DNA/RNA sequence.

Indeed, it has been found that R-loops form *in vivo* when the non-template DNA strand and accordingly the RNA strand are Guanine rich. This was supported by the finding that R-loops form *in vitro* and *in vivo* at G-rich immunoglobulin class switch regions of murine stimulated B-cells **(K. Yu *et al.*, 2003)**. Interestingly, G-rich sequences were found to block transcription *in vitro* through formation of unusually stable R-loops which were speculated to be further complexed with triplex H-DNA formation**.** The formation of R-loops and the associated blockage were dependent on the direction of transcription that they obviously occur when the non-template strand and the RNA transcript are G-rich **(Belotserkovskii *et al.*, 2010)**.

These observations are consistent with the findings that RNA-DNA hybrids with G-rich RNA have more thermodynamic stability compared to the respective DNA duplexes **(Hung *et al.*, 1994; Kankia & Marky, 1999; Lesnik & Freier, 1995; Ratmeyer *et al.*, 1994; Roberts & Crothers, 1992; Sugimoto *et al.*, 1995)**.

Extensive and elegant work has been conducted by Lieber lab to understand how R-loops can efficiently form over G-rich sequences, using a mammalian class switch region as a model for an *in vitro* transcription system. Lieber lab found that this region has many G-clusters of 3-5 guanines, and that reducing these clusters, even without affecting the density of guanines (G-density), sharply decreased R-loop formation over this region **(Roy *et al.*, 2008)**. This evidenced that G-clustering is essentially required at G-rich regions to promote R-loop formation. As shown in Figure 1.7A, it has been shown that G-clusters are required at R-loop initiation zone (RIZ) for the initiation process of R-loop formation while the high G-density at R-loop elongation zone (REZ) is required for maintaining these R-loops **(Roy & Lieber, 2009)**. These findings present a supporting mechanism for the threading-back model of R-loop formation and suggest that R-loops form as a result of kinetic/energetic competition between the RNA and the non-template DNA strand to hybridize with the template DNA **(Chedin, 2016; Roy & Lieber, 2009)**.

Recently, multiple R-loop profiling studies have shown that R-loops accumulate over unmethylated CGI promoters and GC-rich 3' ends of human genes. Moreover, these regions are specifically characterized by the asymmetric distribution of Guanines and Cytosines over the two strands of the DNA. This distribution pattern is called GC skew **(Ginno *et al.*, 2013; Ginno *et al.*, 2012)**. Importantly, poly-A tracts have been also found to accumulate R-loops *in vivo*. It has been shown that poly-A tracts strongly accumulate R-loops in budding yeast, presumably, through disfavoring nucleosome occupancy over these sequences **(Wahba *et al.*, 2016)**. Interestingly, the propensity of a specific sequence to form R-loop *in vivo* may be predicted from the sequence itself. In addition, a case-by-case study may be required to estimate the R-loop-forming potential by a specific sequence. This can be initially indicated by the stability of the RNA-DNA hybrid. For instance, rG-dC hybrids represent the most stable while rU-dA are the least stable hybrid sequences in general and compared to duplex DNA. rC-dG sequence is expected to be less stable than rG-dC, but slightly stable above dC-dG **(Belotserkovskii, Neil, *et al.*, 2013; Belotserkovskii *et al.*, 2018; Lesnik & Freier, 1995)**.

### 1.5.2 Enhancing factors: triplexes and G-quadruplexes

Non-canonical DNA structures such as H-DNA triplexes and G-quadruplexes may enhance R-loop formation by blocking the non-template DNA strand (Figure 1.7B) and preventing its reannealing to the template DNA **(Belotserkovskii *et al.*, 2018)**. G-quartets or G-quadruplexes are four-stranded DNA structures that form when four Guanines in a G-rich sequence interact by Hoogsteen hydrogen bonds and get organized in a planar conformation. Monovalent cations such as $Na^+$ and $K^+$ stabilize these structures **(Bochman *et al.*, 2012; Burge *et al.*, 2006; Maffia *et al.*, 2020)**.

These structures were hypothesized to enhance the formation of R-loops at Ig class switch regions and telomere repeats **(Daniels & Lieber, 1995; Reaban *et al.*, 1994)**. Excitingly, using electron microscopy, G-quadruplexes were demonstrated to form *in vitro* over Ig class switch and telomeric sequences and *in vivo* as a result of transcription of G-rich sequences cloned in a bacterial plasmid **(Duquette *et al.*, 2004)**. The whole structure was called G-loop as it was found to consist of a co-transcriptionally formed R-loop and a folded non-template DNA strand. Nevertheless, treating the transcribed template with RNase H or Replacing GTP with ITP in transcription reaction destroyed the hybrid and the G-quartets. This suggests that G-quartets form as a result of the high stability of the rG-dC RNA-DNA hybrid, but not the reverse **(Duquette *et al.*, 2004)**. Moreover, it was found that performing *in vitro* transcription in the presence of $Li^+$ or $Cs^+$, which destabilizes G-quadruplexes, didn't affect R-loop formation over Ig class switch regions. Based on these findings, it was concluded that G-quadruplexes aren't required for R-loop formation **(Roy *et al.*, 2008)**. These observations evidence that R-loops stabilize G-quartets, yet the reverse situation can't be excluded.

Triplex H-DNA has been also suggested to enhance RNA-DNA hybrid formation. This has been evidenced by persistent hybrid formation over long Friedreich ataxia GAA-TTC triplet-repeat tracts by T7 RNA Pol *in vitro* transcription and by *in vivo* transcription in bacteria. In addition, pausing of T7 RNA Pol was a hallmark linked to RNA-DNA hybrid formation over these long tracts **(Grabczyk *et al.*, 2007)**. Triplex H-DNA formation between the displaced ssDNA and another duplex sequence was suggested as a model explaining the unusual stability of these RNA-DNA hybrids and the stalling of T7 RNA Pol **(Grabczyk *et al.*, 2007)**. The same model was also suggested for stabilizing RNA-DNA hybrid formation during transcription of (GAA)n repeats transformed in yeast. As a homopurine-

homopyrimidine sequence, (GAA)n was suggested to form H-DNA structure. Guanosines and adenosines in a duplex DNA of GAA-CTT repeats may form Hoogsteen hydrogen bonds with either Guanosines and Adenosines or, alternatively, Cytosines and Thymidines, of one DNA strand leaving the second strand to anneal with the RNA transcript and form RNA-DNA hybrid. The whole structure was called H-loop and was suggested to induce repeat contraction and DNA break in an H-loop/replication collision dependent mechanism **(Neil *et al.*, 2018)**.

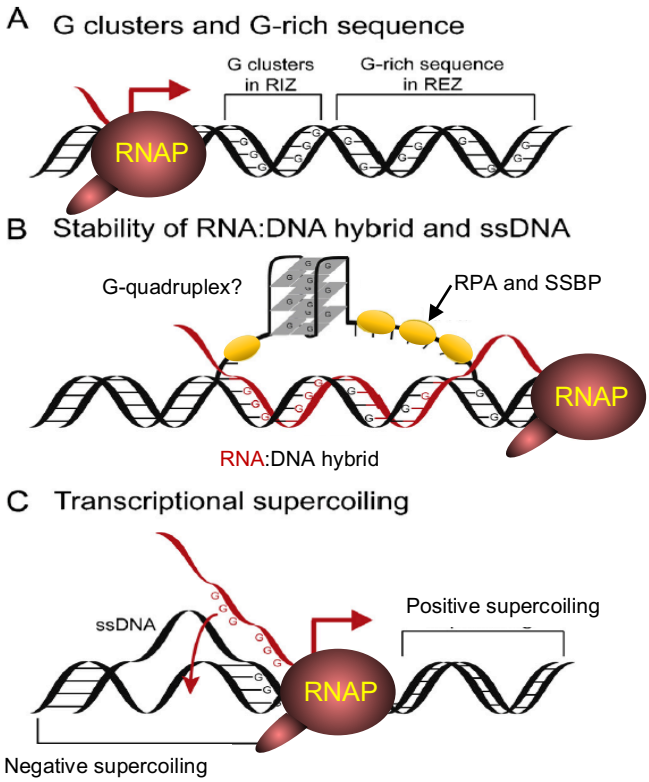



**Figure 1.7. Factors promoting R-loop formation.** Schematic representation for cotranscriptional R-loops and factors promoting their formation. **(A)** G-clusters at R-loop initiation zone (RIZ) trigger R-loop formation wile G-rich sequence at R-loop elongation zone (REZ) is required for R-loop elongation. **(B)** G-quadruplexes and single stranded DNA (ssDNA) binding proteins such as replication protein A (RPA) may stabilize R-loops. **(C)** Negative supercoiling facilitates unwinding of the DNA and annealing of RNA to the template DNA. Adapted from **Hamperl and Cimprich (2014)**.

### 1.5.3 Enhancing factors: DNA single strand break (DNA nick break)

Relevant to the work done by Lieber lab to understand R-loop formation through G-clustering and threading back model (Sections 1.3.1 & 1.5.1), the same lab found that increasing the distance between the promoter (and, respectively, the transcription start site) and R-loop-initiation zone (RIZ or G-clustering region) decreased the formation of R-loops **(Roy *et al.*, 2010)**. This suggested that the presence of a long sequence upstream G-clustering and

R-loop-initiation zone (Figure 1.7A) decreases the propensity of the RNA to anneal back to the DNA strand and, accordingly, suppresses R-loop formation. This was proved by the observation that partial digestion for the 5' end of the RNA transcript during *in vitro* transcription, increased the efficiency of R-loop formation. The conclusion was that the presence of G-clustering region close to the 5' end of the transcript is a requirement for efficient R-loop formation **(Roy *et al.*, 2010)**.

Another outstanding finding was that nicking the non-template DNA strand downstream the promoter increased R-loop formation over the template strand of the nicked compared to the non-nicked DNA, even though transcription over the nicked was much lower than that over non-nicked DNA. Moreover, introducing a nick enhanced R-loop formation over non-G-dense DNA **(Roy *et al.*, 2010)**. This implies that the presence of a nick in the non-template DNA strand may act as an R-loop initiation zone (RIZ), most probably, by disrupting and destabilizing the template/non-template DNA interactions and favoring hybrid formation. DNA nicking and R-loop accumulation were hypothesized as models for recombination mechanism at Ig class switch regions. Interestingly, increasing the distance between the promoter and the nick decreased R-loop formation **(Roy *et al.*, 2010)**.

A study investigating *in vitro* transcription blockage over G-rich homopurine-homopyrimidine sequences due to R-loop formation, found that changing guanine to another base, in DNA sequence, strongly reduced the blockage **(Belotserkovskii *et al.*, 2010)**. The blockage reduction effect was more significant with cytosine or thymine compared to adenine, suggesting that purine sequences enhance R-loop formation. More importantly, introducing a break in the non-template strand adjacent to the poly-G sequence strongly increased the blockage which was much weaker if the nick was in the template strand **(Belotserkovskii, Neil, *et al.*, 2013)**. This presents another *in vitro* evidence that R-loop formation is enhanced by a break in the non-template strand.

### 1.5.4 Enhancing factors: Negative supercoiling.

Elongating RNA Polymerases induce hyper-positive supercoiling stress on the DNA ahead of transcription while leaving the DNA behind in a negative-supercoiled configuration **(L. F. Liu & Wang, 1987; Wu *et al.*, 1988).** Importantly, negative supercoiling and topoisomerases were the key for discovering and studying non-canonical nucleic acid structures such as D-loops **(J. C. Wang, 1974)**, R-loops **(Drolet *et al.*, 1994; Drolet *et al.*,**

**1995)** and H-DNA **(Kohwi & Kohwi-Shigematsu, 1988; Mirkin *et al.*, 1987)**. In fact, there is no non-canonical DNA structure that its formation hasn't been linked to negative supercoiling.

Different *in vitro* and *in vivo* studies investigated the correlation between negative supercoiling and R-loop formation. Loss of topoisomerase I has been shown to induce hyper-negative supercoiling for plasmids, accumulate R-loops *in vivo* and induce lethality phenotype in *E. coli*. The lethality phenotype was rescued with RNase H overexpression **(Massé & Drolet, 1999a, 1999b)**. Furthermore, *E. coli* lacking topoisomerase I (*topA*Δ) were unable to generate full length RNAs at suboptimal temperature. This was solved by RNase H overexpression, implying a role for negative supercoiling in facilitating R-loop formation and hindering RNA synthesis **(Baaklini *et al.*, 2004)**.

Comparing R-loop formation efficiency over linear DNA fragments to that over negatively supercoiled plasmids, during *in vitro* transcription, showed that negatively supercoiled plasmids favor R-loop formation compared to linear DNA fragments. Negatively supercoiled plasmids accumulated R-loops even over non G-rich sequences **(Roy *et al.*, 2010)**. The clear reason is that negative supercoiling facilitates DNA unwinding and strand separation (Figure 1.7C). Notably, the effect of negative supercoiling, separately, as a factor enhancing R-loop formation was found to be stronger than the effect of G-clustering at R-loop initiation zone **(Roy *et al.*, 2010)**. Consistent with these findings, the transcription blockage induced by R-loop formation over G-rich sequences, during *in vitro* transcription, was found to depend on supercoiling. It has been estimated that the transcription blockage effect on supercoiled plasmids is two times stronger compared to that on linear DNA fragments **(Belotserkovskii *et al.*, 2010)**.

An interesting finding linking supercoiling, R-loop formation and RNA Pol I transcription came from a study on *S. Cerevisiae,* which found that deleting Top1 leads to production of truncated ribosomal pre-RNAs, inefficient pre-rRNA synthesis and R-loop accumulation over rDNA repeats, especially, at the 18S rDNA 5' end. Additional loss of Top2 led to severe accumulation of short 18S pre-rRNAs transcripts **(El Hage *et al.*, 2010)**. Interestingly, RNA Pol I piling up could be detected in wild type cells by EM over 18S rDNA. This piling up got increased in *top1*Δ and more detected in *top1*Δ *rnh*Δ strains. These observations were explained by the fact that unresolved negative supercoiling behind RNA Pol

I enhances R-loop formation which blocks transcription at the 5' end of 18S rDNA. Top2 loss increases the positive super-helical stress ahead of RNA Pol I which exacerbates the transcription blockage. Apparently, high transcription rate may have the same effect on DNA supercoiling and inhibit transcription elongation **(El Hage *et al.*, 2010)**.

Altogether, these observations confirm and establish a direct correlation between negative supercoiling and R-loop formation in which, as a general rule, negative supercoiling boosts R-loop formation potential over an arbitrary sequence.

## 1.6 R-loops preventive and resolving surveillance mechanisms

Permanent, or unprogrammed R-loop formation can have deleterious impacts on genome integrity and DNA-related biological processes such as replication and transcription. To control R-loop formation, living cells use different mechanisms to either resolve R-loops or prevent their formation. While R-loop formation is a conserved feature of the genome of all living organisms, R-loop surveillance mechanisms are also conserved among different organisms **(Aguilera & Garcia-Muse, 2012; Garcia-Muse & Aguilera, 2019; Stephan Hamperl & Cimprich, 2014; Santos-Pereira & Aguilera, 2015; Skourti-Stathaki & Proudfoot, 2014)**.

### 1.6.1 Preventive mechanisms: mRNA biogenesis and processing factors

Beside topoisomerases which modulate DNA supercoiling, and chromatin factors such as FACT histone chaperone which maintains nucleosome occupancy and regulates chromatin accessibility, different mRNA biogenesis and processing factors prevent R-loop formation (Figure 1.8). The first evidence implicating mRNA biogenesis factors in preventing R-loop formation was provided by Aguilera lab who revealed that depletion of Hpr1 in *S. Cerevisiae* leads to accumulation of co-transcriptional R-loops **(Huertas & Aguilera, 2003)**. Hpr1 is a component of the THO/TREX, a conserved eukaryotic complex linking mRNA elongation and packaging to the export process (Figure 1.8). The THO complex, which was first characterized to be involved in transcription elongation, consists of Hrp1, Tho2, Mft1 and Thp2 nuclear proteins **(Aguilera, 2002; Chávez *et al.*, 2000)**. The TREX (transcription-export complex) consists of Tex1, Sub2 RNA helicase and Yra1 RNA export factor **(Strässer *et al.*, 2002)**.

Elegant work by Aguilera lab has shown that nascent RNA self-cleavage by engineered hammer head ribozyme, suppressed transcription impairment and hyper-recombination rate

induced by R-loops **(Huertas & Aguilera, 2003)**. Interestingly, RNase H overexpression abolished R-loop formation and hyper-recombination phenotypes, supporting the conclusion that these events are co-transcriptionally correlated **(Huertas & Aguilera, 2003)**. This elegant work provided new insights about the contribution of R-loops to transcription associated recombination (TAR) and genomic instability.

Later, alternative splicing factor/splicing factor 2 (ASF/SF2), an mRNA splicing factor and a member of the Serine Arginine (SR)-rich proteins, was shown to prevent R-loop formation in chicken and human DT40 B cell lines **(X. Li & Manley, 2005)**. loss of ASF/SF2 induced a hypermutation phenotype characterized by DNA double-strand breaks (DSB) and chromosomal arrangements. RNase H overexpression suppressed the hypermutation phenotype, indicating a direct role of ASF/SF2 in inhibiting R-loop formation and evidencing for the involvement of splicing factors in such role in metazoans **(X. Li & Manley, 2005)**.

Another evidence connecting mRNA export machinery to co-transcriptional R-loop prevention came through work on Thp1-Sac3-Sus1-Cdc31 (THSC) complex. THSC, also known as C complex, is known to co-localize and interact with specific nucleoporins at the inner side of the nuclear pore complex (NPC) **(Fischer *et al.*, 2002)**. It consists of the RNA export factor Thp1-Sac3 and the SAGA transcription initiation factor Sus1 **(Rodríguez-Navarro *et al.*, 2004)**. *S. Cerevisiae* mutants of the THSC/TREX-2 exhibited phenotypes comparable to those of the THO/TREX complex mutants. Deleting any of the members of the THSC increased transcription associated hyperrecombination and induced transcription elongation defects. Strikingly, inactivation of the THSC by thp1 deletion induced the mutating activity of human activation-induced cytidine deaminase (AID). This enzyme known to act on single stranded DNA at Ig class switch regions, which strongly suggests the accumulation of R-loops as a result of THSC inactivation **(González-Aguilera *et al.*, 2008)**.

A recent study screening the genome-wide distribution of THOC7, a member of human THO complex, found that it accumulates at repetitive regions of the human genome and overlaps with R-loop positive loci and, also, Ser2-phosphorylated RNA Pol II, a mark of transcription elongation. Depletion of THOC5, an RNA binding member of the human THO, led to accumulation of telomeric aberrations in human cells. These findings suggest a role for transcription elongation and RNA export factors in preventing genomic stability in human by suppressing R-loop formation **(Katahira *et al.*, 2020)**.

These observations confirm that mRNA processing and export factors play a crucial role in suppressing R-loop formation. However, the exact mechanism behind isn't clear. It has been suggested that instant co-transcriptional packaging of mRNA into ribonucleoprotein complexes, completely sequesters the RNA and prevents hybridization with DNA (Figure 1.8). Consistent with this suggestion, defects in mRNA export can increase the half-life time of mRNAs in the nucleus, or, by another meaning, retain mRNAs in the nucleus which increases the chance of their reannealing to the complementary DNA. It can't be excluded that these factors indirectly prevent R-loop formation by maintaining the proper elongation rate of RNA Pol which if impaired, or arrested would lead to RNA-DNA hybridization **(Aguilera & Garcia-Muse, 2012; Garcia-Muse & Aguilera, 2019; Stephan Hamperl & Cimprich, 2014; Santos-Pereira & Aguilera, 2015; Skourti-Stathaki & Proudfoot, 2014)**.

In prokaryotes, it has been suggested that coupling translation to transcription prevents R-loop formation by immediately loading the nascent transcript into a ribonucleoprotein complex that would prevent reannealing of the RNA to the template DNA strand (Figure 1.8). Active translation along with other transcription elongation and termination factors work together to prevent RNA backtracking of arrested RNA Pol in *E. coli* **(Dutta *et al.*, 2011; Gowrishankar & Harinarayanan, 2004; Gowrishankar *et al.*, 2013; Nudler, 2012).**

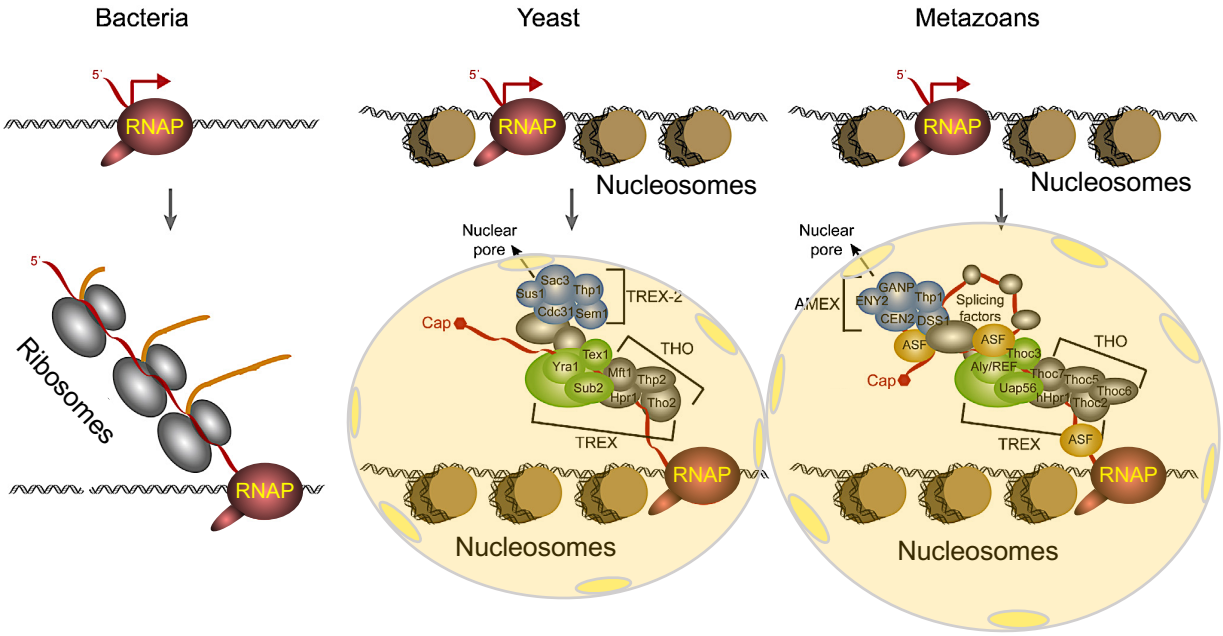



**Figure 1.8. Factors preventing R-loop formation.** In prokaryotes, immediate loading for the nascent transcript into a ribonucleoprotein complex prevents R-loop formation. In yeast and mammalian cells, DNA is wrapped around nucleosome particles (in brown) which may prevent RNA invasion and hybrid formation. In addition, different RNA processing, splicing and export factors form ribonucleoprotein complexes that prevent the RNA form threading back. Adapted from **Hamperl and Cimprich (2014)**.

### 1.6.2 Preventive mechanisms: mRNA surveillance and poly adenylation factors

Interestingly, the mRNA surveillance Machinery has been found to suppress R-loop formation. Loss of Pab2 (also known as Trf4), the polyadenylation (pA) polymerase of the TRAMP complex in *S. cerevisiae*, led to accumulation of R-loops and manifested a hyper-recombination phenotype. This hyper-reombination phenotype got suppressed by RNase H overexpression and, on the contrary, exacerbated by human AID **(Gavalda *et al.*, 2013)**. The TRAMP complex is a polyadenylation complex consisting of Trf4, Air2 and Mtr4 proteins. It's known to be involved in RNA surveillance by polyadenylating RNAs and promoting their degradation by the exosome **(LaCava *et al.*, 2005; Vanácová *et al.*, 2005; Wyers *et al.*, 2005)**.

The exosome has also been implicated in suppression of R-loop formation. Depletion of the RNA exonuclease components of the exosome such as EXOSC3 and EXOSC10 (exoribonucleases exosome component 3 and 10), the homologues of yeast Rrp40 and Rrp6 respectively, led to accumulation of R-loops, especially, over divergent enhancer RNA (eRNA)-expressing regions in mouse embryonic stem cells **(Pefanis *et al.*, 2015)**.

The fact that the different mechanisms involved in prevention of R-loops are implicated in different steps of RNA biogenesis and contribute variably to genomic stability implies that these R-loops may be different in structure and function as well. This also implies that R-loop formation and prevention depends on the structure and fate of the RNA.

### 1.6.3 Preventive mechanisms: Recombination and DNA repair machinery

R-loops are known to prevent the progression of replication fork (RF), mainly, during the S-G2 phase of the cell cycle which induces DNA double strand break (DSB) damage and genomic instability, a hallmark of cancer **(Crossley *et al.*, 2019; Garcia-Muse & Aguilera, 2019; Gómez-González & Aguilera, 2019)**. Multiple pathways and mechanisms act to maintain the RF progression and avoid its collapse. Of these, are the DNA repair mechanism mediated by the Fanconi anemia complex factors which are implicated in processing DNA lesions that block replication and in repairing inter-strand crosslinks (ICLs).

The role of Fanconi anemia complex in suppressing R-loop formation was first revealed in mammalian cells depleted of BRCA2. BRCA2 was found to suppress R-loop formation and bind with PCID2, a member of the TREX-2 mRNA export factor **(Bhatia *et al.*, 2014)**. This

finding was confirmed and extended by other studies showing that depletion of BRCA1 **(Hatchi *et al.*, 2015)**, FANCM **(Schwab *et al.*, 2015)** or FANCD2/FANCA **(Garcia-Rubio *et al.*, 2015; Liang *et al.*, 2019; Madireddy *et al.*, 2016; Okamoto *et al.*, 2019)**, led to accumulation of R-loops and iduced DNA damage.

## 1.7 Resolving factors: RNase H enzymes and RNA-DNA helicases

### 1.7.1 Resolving factors: RNase H enzymes

Ribonuclease H (RNase H) enzymes are dedicated for the degradation of the RNA moiety of the RNA-DNA hybrid in a sequence independent manner **(Cerritelli & Crouch, 2009)**. Most organisms, including viruses, have at least one copy of these enzymes which are classified based on sequence conservation and substrate preference into two main types. These are type 1 referred to as RNase H1 in eukaryotes and RNase HI in prokaryote, and type 2 referred to as RNase H2 and RNase HII in eukaryotes and prokaryotes, respectively **(Cerritelli & Crouch, 2009)**. Eukaryotic RNase H1 is larger and more complex than its prokaryotic counterpart, but they are both monomeric and similar in structure. Eukaryotic RNase H2 has three subunits, RNase H2A, H2B and H2C while the prokaryotic enzyme is monomeric in structure. During early embryonic development in mouse, RNase H1 is required for mitochondrial DNA replication **(Cerritelli & Crouch, 2009)**.

Eukaryotic RNase H1 has two highly conserved domains at the N- and C-termini and a variable connecting region. The N-terminal end contains the hybrid-binding domain (HBD) which imparts substrate specificity and processivity. It has 25 times more affinity toward RNA-DNA hybrid compared to dsRNA of the same sequence. While RNase H1 of higher eukaryotes possesses one HBD, *S. cerevisiae* have two domains, the second one is thought to increase the binding with the RNA-DNA hybrid. The C-terminus contains the enzymatic domain, also known as RNase H domain. Both the HBD and RNase H domains are linked by a connection domain that confers flexibility to the enzyme. RNase H1 of higher eukaryotes contains a mitochondrial targeting sequence (MTS) that is expressed during early development stages to target RNase H1 to the mitochondria where it's essential for mitochondrial DNA replication **(Cerritelli & Crouch, 2009)**.

Each of the type 1 and type 2 RNase H domains exhibits substrate sequence preferences. Type 1 RNase H can't cleave hybrid RNA that is shorter than four successive ribonucleotides. On the other hand, type 2 RNase H can cleave even a single ribonucleotide inserted in dsDNA sequence (Figure 1.9). Depletion of RNase H, especially, RNase H1 leads to accumulation of R-loops, DNA damage and genomic instability. Recombinant RNase H is used for *in vitro* treatment of hybrids for controlling R-loop mapping experiments. Endogenous RNase H is also over-expressed *in vivo* to check the sensitivity of mapped signals and confirm the reliability of the data.

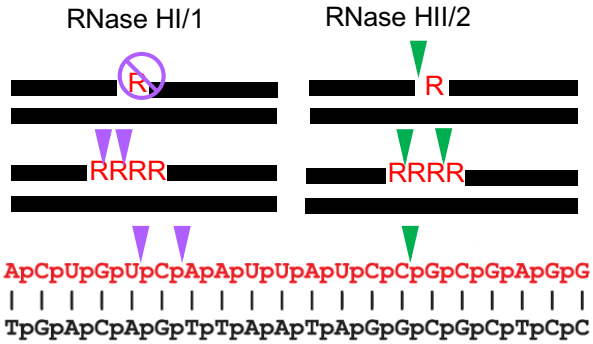



**Figure 1.9. Different cleavage patterns of both RNase H1 and RNase H2 for ribonucleotides in DNA.** Only RNase H2 can cleave a single ribonucleotide embedded in dsDNA (top). In addition, RNase H1 and RNase H2 differently process successive ribonucleotides embedded in dsDNA (middle), or in RNA-DNA hybrid (bottom). Adapted from **Critelli and Crouch (2009)**.

### 1.7.2 Resolving factors: RNA-DNA helicases

RNA-DNA helicases are the second group of enzymes known to resolve R-loops (Figure 1.10). Recently, an increasing number of RNA-dependent ATP-ases have been found to be able to unwind the RNA-DNA hybrids *in vitro* and to accumulate R-loops if depleted *in vivo*. In *E. Coli,* the DNA translocase RecG was found to target a variety of substrates including holiday junctions and R-loops **(Rudolph *et al.*, 2010)**. Another example from bacteria is the transcription termination factor Rho which was found to have a DNA-RNA helicase activity. Interestingly, inducing Rho-dependent termination deficiency was found to accumulate R-loops suggesting that it has R-loop surveillance function and that it protects the genome from R-loop accumulation **(Leela *et al.*, 2013)**.

In eukaryotes, different helicases have been identified to resolve R-loops. These include human RNA-DNA helicases SETX and FANCM (homologs of yeast Sen1 and Pif1 family

helicases) and the human DHX9/RNA helicase A **(Garcia-Muse & Aguilera, 2019; Stephan Hamperl & Cimprich, 2014)**. Senataxin 1 (Sen1) was first identified as an RNA Pol II-associated transcription termination factor and as a member of the Sen1-Nrd1-Sab3 complex required for termination of small RNAs that are not poly-adenylated **(Hazelbaker *et al.*, 2013)**. Interestingly, Sen1 has been shown to resolve co-transcriptionally formed RNA-DNA hybrids and prevent transcription-associated genome instability **(Mischo *et al.*, 2011)**. Moreover, Sen1 has been found to resolve R-loops formed over terminal transcription pause sites and to facilitate Xrn2-mediated transcription termination **(Skourti-Stathaki *et al.*, 2011)**. Although it has been known as an RNA Pol II transcription termination factor, intriguingly, Sen1 has been lately found to promote transcription termination of RNA Pol III. This role was found to be independent from R-loops **(Rivosecchi *et al.*, 2019)**.

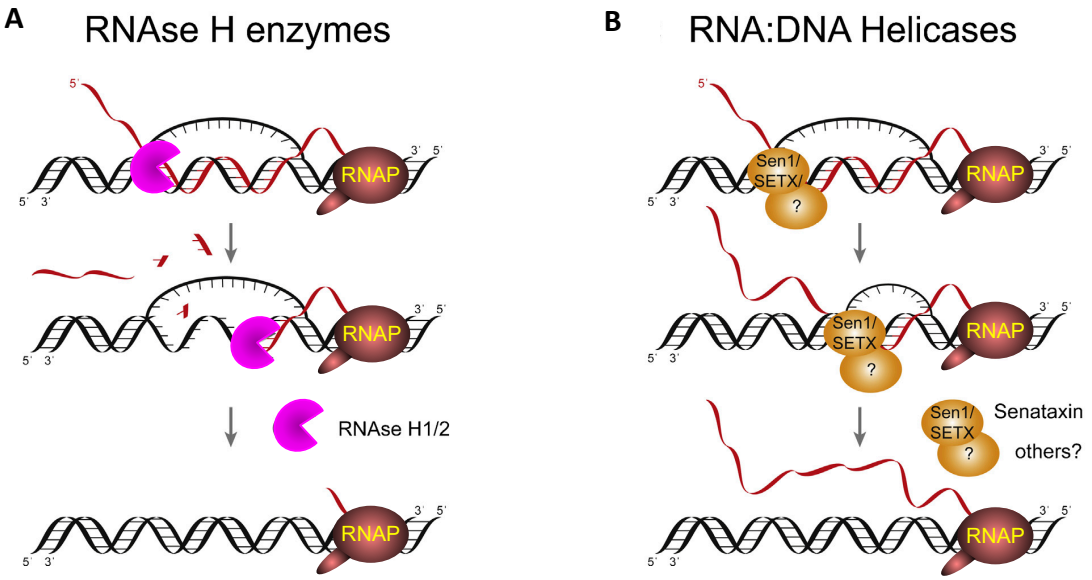



**Figure 1.10. Major R-loop-resolving enzymes. (A)** RNase H enzymes cleave the RNA strand of the RNA-DNA hybrid. **(B)** RNA-DNA helicases such as senataxin and its homologs, Sen1 or SETX, unwind the RNA-DNA hybrid. Adapted from **Hamperl and Cimprich (2014)**.

## 1.8 Regulatory roles of R-loops

Since their discovery *in vivo* in *E. coli* **(Drolet *et al.*, 1994; Drolet *et al.*, 1995)**, R-loops have been considered for long as either rare transcriptional by-products or hazardous structures that threaten genomic stability of different organisms **(Aguilera & Garcia-Muse, 2012)**. Recently, R-loops have been found to be more abundant than previously thought and to form frequently over the genome of different organisms. It has been estimated that from 5 to 10% of the eukaryotic genome has the ability to form R-loops.

The number of reports about R-loops has recently exploded as R-loops have gained much interest as potential regulators of gene expression and functionally relevant structures implicated in many nuclear processes and linked to many diseases. For instance, R-loops have been shown to be required for initiation of mitochondrial replication **(Lee & Clayton, 1998)** and immunoglobulin class switch recombination in B-lymphocytes **(K. Yu *et al.*, 2003)**. Excitingly, R-loops have been shown to form over DNA break ends **(Cohen *et al.*, 2018; Ohle *et al.*, 2016)** and suggested to play a role in DSB repair. Recently, R-loops have been implicated in telomere repair and maintenance and in preventing premature senescence **(Balk *et al.*, 2013; Graf *et al.*, 2017; Perez-Martinez *et al.*, 2020)**. For the next part, I will focus on revealing the different roles of R-loops in transcription regulation in conjunction with antisense transcription.

### 1.8.1 Regulatory roles of R-loops in transcription activation and repression

In general, R-loop profiling studies have shown that R-loops accumulate at hypomethylated CpG island promoters with high GC-skew in mammalian cells **(Ginno *et al.*, 2013; Ginno *et al.*, 2012; Sanz *et al.*, 2016)**. Consistently, these promoters are rich in histone marks of active transcription such as H3K4me3, H4K20me1 and H3K79me2 (Figure 1.11A). These R-loops are believed to prevent the action of DNA methyl-transferases and prevent gene silencing **(Grunseich *et al.*, 2018)**. Interestingly, R-loops have been shown to induce the binding of specific proteins such as GADD45A protein that recruits the TET1 demethylase for local DNA demethylation **(Arab *et al.*, 2019)**.

R-loops have been found to regulate gene expression in variable context-dependent ways. In fact, rules governing R-loop behavior regarding activation or repression of transcription aren't clear. It's thought that these processes may be sequence- and chromatin-context dependent **(Crossley *et al.*, 2019)**. In particular cases, R-loop-mediated regulation is dependent

on antisense ncRNA. In human colon adenocarcinoma cell lines, head-to-head antisense transcription at vimentin (*VIM*) gene induces R-loop formation which is thought to activate antisense transcription, decrease nucleosome occupancy and maintain chromatin openness (Figure 1.11B). These R-loops enhance the binding of nuclear factor-κB (NF-κB), a transcription activation factor that activates *VIM* gene **(Boque-Sastre *et al.*, 2015)**.

In *Arabidopsis*, antisense R-loop formation was found to regulate gene expression at the FLOWERING LOCUS (FLC) in a different way. At cold conditions, antisense transcription from COOLAIR promoter epigenetically silences the upstream FLC sense promoter. Persistent R-loop formation over COOLAIR promoter inhibits transcription from this promoter. This maintains the sense transcription activation of the upstream FlC promoter (Figure 1.11C). Formation of R-loops recruits the ssDNA-binding AtNDX protein which stabilizes R-loops and prevents their resolution **(Sun *et al.*, 2013)**.

In mouse brain cells, specifically, at ubiquitin ligase E3a (Ube3a) gene, R-loop formation by antisense ncRNA has been shown to be involved in transcription regulation (Figure 1.11D). Antisense ncRNA transcription for the small nuclear RNA Snord116, downstream ubiquitin ligase, produces a readthrough asRNA (Ube3a-ATS) that silences the paternal copy of Ube3a gene. Treating with Topotecan, a topoisomerase 1 inhibitor, reactivates paternal Ube3a gene by suppressing Ube3a-ATS transcription (Figure 1.11D). This process was found to involve the stabilization of R-loop formation by repetitive elements of paternal Snord116, which stalls progression of Pol II to Ube3a-ATS and induces chromatin de-condensation at this locus **(Powell *et al.*, 2013)**.

At the promoter of tumor suppressor factor TCF21, antisense transcription of TARID long non-coding RNA (lncRNA) induces R-loop formation at TCF21 promoter. R-loop formation recruits GADD45A protein and brings TET1 demethylase which demethylates the promoter and activates transcription (Figure 1.11A) **(Arab *et al.*, 2019)**. In contrast to this situation, R-loops were found to enhance the repression of specific regulatory genes in mouse embryonic stem cells through the Polycomb repressor proteins **(Skourti-Stathaki *et al.*, 2019)**.

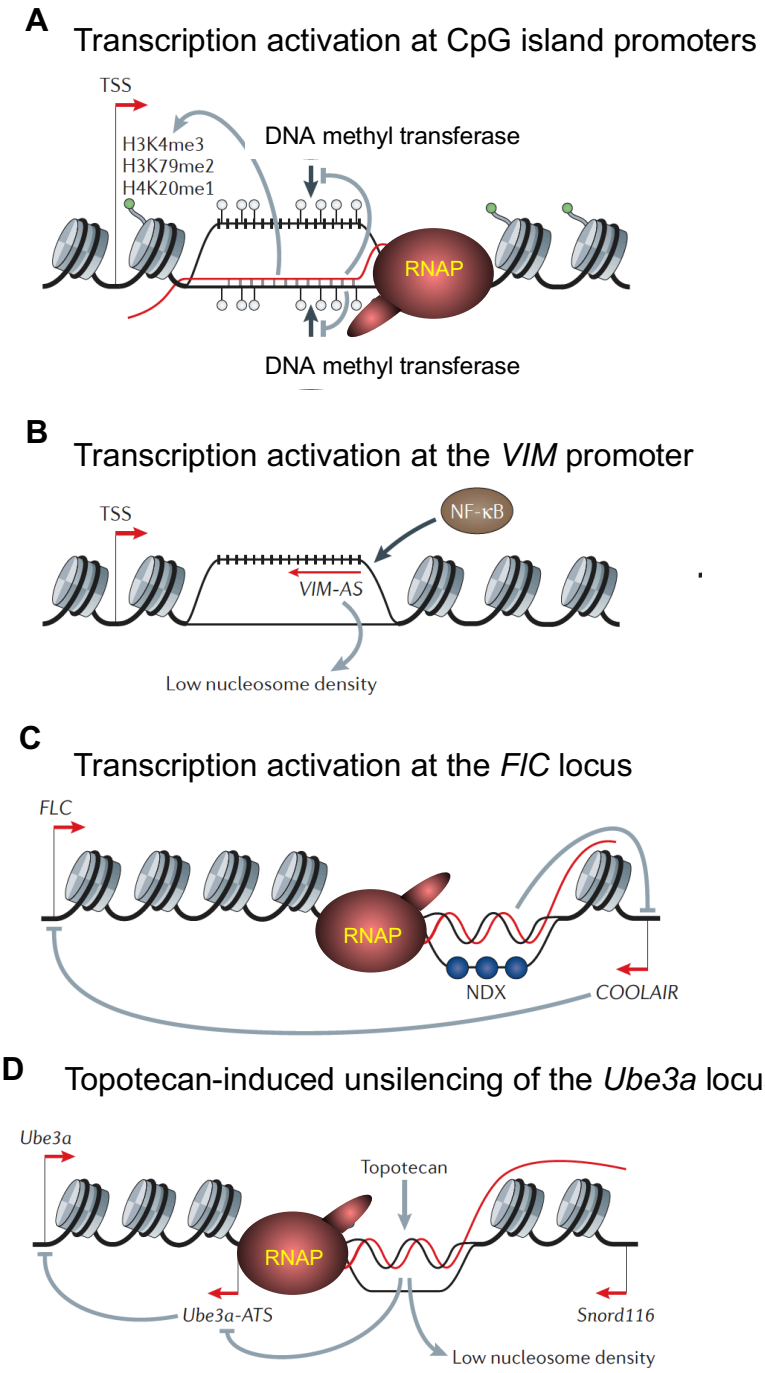



**Figure 1.11. Representative examples for R-loops involved in transcription regulation. (A)** R-loops prevent the action of DNA methyl-transferases and prevent gene silencing. **(B)** head-to-head antisense transcription at vimentin (*VIM*) gene induces R-loop formation and *VIM* transcription activation **(C)** R-loop stabilization over COOLAIR promoter inhibits antisense transcription and maintains FLC promoter activation **(D)** R-loop formation by SNORD116 antisense ncRNA has been shown to be involved in transcription activation of ubiquitin ligase E3a (Ube3a) gene**.** See text for more details. Adapted from **Santos-Pereira and Aguilera (2015)**.

## 1.9 R-loop mapping and detection methods

All the above R-loop related findings have been enabled by advances in R-loop mapping and detection technologies. Several techniques and approaches have been recently introduced to study and map R-loops and address many relevant questions **(Chédin *et al.*, 2021; Crossley *et al.*, 2019; Garcia-Muse & Aguilera, 2019; Halasz *et al.*, 2017; Vanoosthuyse, 2018)**. The vast majority of these methods depend on direct pulldown of R-loops using the S9.6 antibody and next generation sequencing (Figure 1.12 and Table 1.1).

### 1.9.1 Anti-hybrid S9.6 antibody

S9.6 is a monoclonal antibody raised against RNA-DNA hybrids **(Boguslawski *et al.*, 1986)**. It can identify RNA-DNA hybrids as short as 6-base-pair long, making dependent techniques the most reliable for detection of R-loops **(F. Konig *et al.*, 2017)**. However, this antibody has been found to have binding affinity toward dsRNA, but to a lower extent compared to RNA-DNA hybrids **(Hartono *et al.*, 2018; Phillips *et al.*, 2013; Z. Z. Zhang *et al.*, 2015)**. In addition, S9.6 shows high bias in binding affinity based on nucleotide sequence of RNA-DNA hybrids **(F. Konig *et al.*, 2017)**. This antibody has been widely used for detecting R-loops using dot blotting, immunofluorescence and immunoprecipitation.

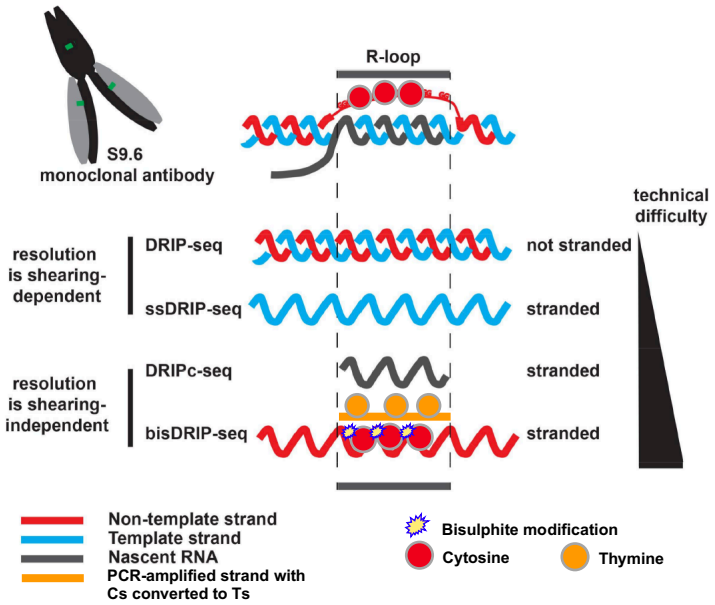



**Figure 1.12. Different DRIP-like methods used for genome-wide mapping of R-loops.** DRIP maps dsDNA, ssDRIP captures ssDNA and maps DNA strand of RNA-DNA hybrid while DRIPc captures RNA strand of R-loops. bisDRIP uses bisulphite conversion and maps converted cytosines in free ssDNA of R-loops. Adapted from **Vanoosthuyse (2018)**.

### 1.9.2 S9.6-based DRIP and DRIP-like methods: DNA-based methods

#### *1.9.2.1 DRIP and DRIP-ChIP*

S9.6 was first used to map RNA-DNA hybrids at specific genomic loci in budding yeast **(El Hage *et al.*, 2010)** and human cells **(Skourti-Stathaki *et al.*, 2011)**. Later, S9.6 was used to map genome-wide R-loops in human cells by a technique termed DNA-RNA immunoprecipitation, or DRIP. The mapped R-loops were confirmed by sensitivity to RNase H treatment **(Ginno *et al.*, 2012)**. This technique is based on the extraction of total nucleic acids that are later treated with restriction enzymes cocktail for fragmentation before the immunoprecipitation step. Finally, the DNA is sequenced, and reads are mapped to the genome. Although it has been very helpful, the DRIP, in its original form, had multiple caveats such as low resolution and lack of directionality.

Different changes have been applied to the original protocol to improve different aspects. Importantly, different sequencing methods have been introduced to enhance resolution and directionality (Figure 1.12 & Table 1.1). Another variant of the DRIP called DRIP-ChIP was used to map R-loops in budding yeast. This version incorporated crosslinking and the use of chromatin preparation and sonication instead of nucleic acid extract and restriction enzymes treatment, respectively **(El Hage *et al.*, 2014)**. Different from the DRIP, the DRIP-ChIP have enabled capturing *in situ* R-loops and enhanced the resolution, but still couldn't solve the directionality problem.

#### *1.9.2.2 S1-DRIP and ChIP-exo*

S1-DRIP is a method that adopted the use of S1 nuclease to digest the free ssDNA of R-loops in the nucleic extract before sonication, in order to map only the DNA strand of the hybrid and enhance directionality of reads. However, after DNA sequencing, the mapped reads were found to be not stranded **(Wahba *et al.*, 2016)**. The high-resolution chromatin immunoprecipitation-exonuclease (ChIP-exo) technique has been also used to map R-loops in fission yeast **(Ohle *et al.*, 2016)**. Although it produced high-resolution R-loop signals, the directionality of this methods hasn't been verified.

#### *1.9.2.3 ssDRIP*

The best-known directional and DNA-based DRIP-like method is the ssDRIP (single stranded DNA-based DRIP) method that was applied to *Arabidopsis.* It incorporated a single stranded DNA library preparation method for which the extracted DNA was ligated to single stranded adaptors **(Xu *et al.*, 2017)**. This method delivered directional sequencing, but its resolution is still dependent on fragment length of DNA.

### 1.9.3 S9.6-based DRIP and DRIP-like methods: RNA-based methods

#### *1.9.3.1 RDIP*

All the above-mentioned methods are based on processing DNA as a template for sequencing. A different class of the DRIP methods, ending by library preparation and sequencing for the RNA strand, was developed to deliver stranded mapping with higher resolution compared to the DNA-based DRIP. The first directional method for R-loop mapping is the RDIP **(Nadel *et al.*, 2015)** which uses total nucleic acids as immunoprecipitation input. After sonication, nucleic acid extract is treated with RNase I to digest free ssRNA before immunoprecipitation. The IPed RNA-DNA hybrids are prepared for directional mapping following directional RNA library preparation workflow, but, importantly, the hybrids are kept intact without eliminating the DNA strand. Instead, the RNA strand is nicked by RNase H to produce short primers required for extension and second DNA strand generation using dUTPs instead of dTTPs. After adaptor ligation, the dUTP-labelled DNA strand is digested leaving directional information in the form of the original R-loop-forming DNA strand.

#### *1.9.3.2 DRIPc*

An improvement for RDIP, is DRIPc (DRIP and cDNA synthesis), an RNA-based DRIP, which depends completely on RNA processing in absence of DNA stand. After DNase treatment to digest the DNA from immunoprecipitated hybrids, RNA is reverse transcribed for cDNA synthesis followed by second strand DNA generation using dUTPs to be digested after adaptor ligation **(Sanz *et al.*, 2016)**. Both the RDIP and DRIPc enabled directional and relatively higher resolution mapping for R-loops. However, working with RNA makes these two methods more challenging than the DNA-based DRIP methods as this may affect the sensitivity of the mapping and lead to underrepresentation of some positive loci. Another complication that adds to the RDIP and DRIPc is that the S9.6 antibody has affinity to dsRNA

which may affect the results. This obligates the use of specific nucleases to get rid of these structures.

### 1.9.4 RNase H-based R-loop mapping methods

#### *1.9.4.1 DRIVE-seq*

Making use of the highly conserved RNase H enzymes which can identify and cleave the RNA part of RNA-DNA hybrids in different organisms, some methods have been developed to map R-loops. These methods depend on the use of catalytically inactive RNase H enzyme that can identify and bind RNA-DNA hybrids but not resolve them. DRIVE-seq (DNA-RNA *in vitro* enrichment), the earliest version of these methods, used mutant human RNase H1 tagged with maltose binding domain (MBD). This tagged RNase H was exogenously expressed and purified from bacteria **(Ginno *et al.*, 2012)**. DRIVE-seq method depends on nucleic acid extraction then fragmentation by restriction enzymes, followed by R-loop enrichment using purified MBD-RNase H to be pulled down using amylose beads for DNA extraction and sequencing. This method captures fewer R-loops compared to the DRIP-like methods.

#### *1.9.4.2 R-ChIP-seq and RR-ChIP-seq*

DRIVE-seq method was improved by the use of catalytically-deficient endogenous fission yeast RNase H1 (Rnh1) for binding native *in vivo* R-loops. After chromatin preparation, R-loops were indirectly enriched by chromatin immunoprecipitation using the tagged mutant Rnh1 as a bait. This strategy was first used to study enrichment of R-loops over specific genes such as RNA Pol III genes using qPCR **(Legros *et al.*, 2014)**. A similar approach called R-ChIP was adopted for genome-wide mapping of R-loops in human cells **(L. Chen *et al.*, 2017)**. Similar to the improved DRIVE-seq, R-ChIP method requires the stable expression of exogenous, tagged and mutant RNase H1. R-ChIP is a directional R-loop mapping method that adopts random extension using a primer with a linker as a single stranded DNA-capturing method. A catalytically and binding-deficient RNase H1-expresssing strain is used as a negative control. Although this method presented stranded signals, it identified fewer R-loops compared to the DRIP and DRIPc. Remarkably, R-ChIP signals mapped mainly to promoter regions different from DRIP and DRIPc which identified R-loops at promoters and terminators. Another similar method, called RR-ChIP, has been recently introduced to capture the RNA strand instead of DNA strand of RNA-DNA hybrids. Similar to R-ChIP, RR-ChIP identified R-loops mainly at promoters of mammalian cells **(Tan-Wong *et al.*, 2019)**.

**Table 1.1. List of all methods used for genome-wide R-loop mapping and technical differences among them.**
Resolution of some methods is questioned (resolution?) as it depends on fragment size of nucleic acids.

| Method | Input material | Shearing method | Template sequenced | Controls used | Special enzyme | Strengths | Weaknesses | Reference |
|---|---|---|---|---|---|---|---|---|
| **DRIP** | Nucleic acids | Enzymatic digestion | DNA | RNase H treatment | No | Robust | Not *in situ*<br>Not directional<br>Low resolution | **(Ginno *et al.*, 2012)** |
| **ChIP-DRIP** | FA-fixed Chromatin | Sonication | DNA | RNase H treatment | No | Robust<br>*in situ* | Not directional<br>Low resolution | **(El Hage *et al.*, 2014)** |
| **ChIP-exo** | Chromatin prep | Sonication | DNA | No | Multiple enzymes | High resolution<br>*In situ* | Laborious<br>Long & complicated<br>Directionality not confirmed | **(Ohle *et al.*, 2016)** |
| **S1-DRIP** | Nucleic acids | Sonication | DNA | RNase H treatment | S1 nuclease & RNase A | Quantitative | Not *in situ*<br>Not directional<br>Resolution? | **(Wahba *et al.*, 2016)** |
| **ssDRIP** | Nucleic acids | Sonication or enzymatic | Adaptor ligation ssDNA | RNase H treatment | No | Directional | Not *in situ*<br>Resolution? | **(Xu *et al.*, 2017)** |
| **RDIP** | Nucleic acids | Sonication | priming of DNA by RNA strand | No | RNase I | Directional | Not *in situ*<br>Resolution? | **(Nadel *et al.*, 2015)** |
| **DRIPc** | Nucleic acids | Enzymatic | RNA | RNase H treatment | RNase A | Directional | Not *in situ*<br>Affinity of S9.6 for dsRNA<br>Resolution? | **(Sanz *et al.*, 2016)** |
| **Yeast DRIPc** | Nucleic acids | Enzymatic | RNA | RNase H treatment | RNase III | Directional | Not *in situ*<br>Affinity of S9.6 for dsRNA<br>Resolution? | **(Hartono *et al.*, 2018)** |
| **qDRIP** | Nucleic acids | Sonication | ssDNA ligation | RNase H treatment | No | Directional<br>Quantitative | Not *in situ*<br>Multiple spike-ins may be required<br>Resolution? | **(Crossley *et al.*, 2020)** |
| **Drive-seq** | Nucleic acids | Enzymatic | DNA | RNase H treatment | Enriching with *in vitro* RNase H1 | Confirmative method | Not *in situ*<br>Low resolution<br>Not directional | **(Ginno *et al.*, 2012)** |
| **R-ChIP** | FA-fixed Chromatin | Sonication | Random priming of ssDNA | Binding-deficient RNase H1 | No | Directional<br>Maps native dynamic R-loops | Requires stable expression of mutant RNase H<br>Maps few R-loops<br>Resolution? | **(L. Chen *et al.*, 2017)** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **RR-ChIP** | FA-fixed Chromatin | Sonication | Directional RNA | Binding-deficient RNase H1 or RNase H1 OE | No | Directional<br>Maps native dynamic R-loops | Requires stable expression of mutant RNase H<br>Maps few R-loops<br>Resolution? | **(Tan-Wong *et al.*, 2019)** |
| **Bis-DRIP** | Nucleic acids | Enzymatic | Sonicated DNA | RNase H treatment | Sodium bisulphite | Directional<br>Improved resolution | Requires many replicates<br>Resolution is dependent on cytosines | **(Dumelie & Jaffrey, 2017)** |
| **SMRF** | Nucleic acids | Enzymatic | Amplified DNA | RNase H treatment | Sodium bisulphite | High resolution<br>Directional<br>Long read sequencing<br>Ultra-high coverage. | Gene-specific<br>Requires multiple PCRs with target-specific primer | **(Malig *et al.*, 2020)** |
| **CUT&RUN** | Intact nuclei | MNase | DNA | IgG control | No | Captures native R-loops<br>Short time | Captures few R-loops<br>Generation of stable cell lines<br>Cloning and purification for MNase complex<br>Not directional<br>Low resolution | **(Yan *et al.*, 2019)** |
| **MapR** | Intact nuclei | MNase | DNA | GST-MNase without RNase H, RNase H treatment | No | Low input<br>Shortens time<br>Captures native R-loops | Cloning and purification for RNase H-MNase complex<br>Not directional<br>Low esolution | **(Yan *et al.*, 2019)** |
| **CUT&Tag** | Intact cells/nuclei | Tn5 transposase | DNA | RNase H<br>RNase A | RNase A | High sensitivity<br>Relatively high resolution<br>Maps native dynamic R-loops | Cloning and purification for biosensor complex<br>Multiple Ab incubations<br>Cloning and purification for Tn5 transposase<br>Tn5 affinity to accessible DNA<br>Not directional | **(K. Wang *et al.*, 2021)** |

### 1.9.5 Bisulphite conversion methods

#### *1.9.5.1 Bisulphite DRIP-seq*

Under non-denaturing conditions, sodium bisulphite converts non-methylated cytosines in unpaired ssDNA into uracil. This uracil is complemented during PCR by adenine (instead of guanine) to finally replace C-G by T-A base pairs. Bisulphite conversion has been used for directional R-loop mapping through identifying the displaced ssDNA with converted cytosines. Only cytosines in the displaced DNA strand will be converted, while those in in the other strand will be protected by RNA-DNA hybrid formation **(Dumelie & Jaffrey, 2017; K. Yu *et al.*, 2003)**. In some experiments, bisulphite treatment was conducted during cell lysis and in presence of SDS to convert cytosines and avoid expected R-loop resolution or DNA reannealing events that may occur during lysis **(Dumelie & Jaffrey, 2017)**. In addition, an S9.6 enrichment step for R-loops can be conducted after bisulphite conversion and DNA fragmentation required for generating short fragments and eliminating the need for deep high sequencing **(Dumelie & Jaffrey, 2017)**

A great advantage of bisulphite conversion methods is that they allow R-loop mapping at near single nucleotide resolution for GC-rich sequences. However, resolution is dependent on the presence of Cs in the ssDNA of an R-loop. The biggest caveat for this method is that it can capture ssDNA at genomic regions not involved in R-loop formation.

#### *1.9.5.2 SMRF and BacBio deep sequencing*

SMRF or single molecule R-loop foot-printing is a bisulphite conversion dependent method. It adopts long-read single molecule BacBio deep sequencing for a single-nucleotide resolution and directional R-loop mapping. Importantly, this method confirmed the compatibility and results of previous DRIP methods **(Malig *et al.*, 2020)**. A limitation for this method is that the recovery of R-loop molecules for a specific region depends on frequency of R-loop formation. This means PCR amplification may be required to enrich R-loop-forming sequences of some target regions, using gene-specific primers. Alternatively, An S9.6 enrichment step may be included **(Malig *et al.*, 2020)**.

### 1.9.6 Targeted nuclease methods for mapping R-loops

#### *1.9.6.1 CUT&RUN*

Making use of the CUT&RUN approach **(Skene *et al.*, 2018; Skene & Henikoff, 2017)**, two techniques were developed to map genome-wide native R-loops in a relatively short time **(Yan *et al.*, 2019)**. The first method used the default CUT&RUN principle for cells engineered to exogenously express mutant RNase H1 tagged with FLAG epitope. After permeabilization for nuclei, tagged RNase H is then targeted *in vivo* by anti-FLAG antibody which is finally captured by protein A-MNase chimeric protein complex. MNase cleaves nucleic acids in the boundaries of RNase H/R-loops while the RNase H bait (binding R-loops), Protein A-MNase and genomic fragments are released outside the nucleus for further processing.

#### *1.9.6.2 MapR*

The second method, called mapR, targets R-loops directly with an extracellulary produced complex consisting of catalytically inactive RNase H1 fused to MNase. As a negative control, a GST-tagged MNase without RNase H was used. MapR follows a workflow that is similar to CUT&RUN. A great advantage for this method is that it eliminated the need for generating stable cell lines expressing endogenous RNase H. In addition, it has been shown to be efficient with low-input material **(Yan *et al.*, 2019)**. However, it may suffer from low resolution and lack of directionality.

#### *1.9.6.3 Biosensor and CUT&Tag methods*

A new strategy has been recently developed to map R-loops using the hybrid binding domain (HBD) of RNase H1 instead of the full enzyme. This was suggested to overcome the caveats of the above RNase H-based and nuclease targeting methods, especially the need for generating stable cell lines expressing mutant RNase H1 and the inefficient recognition and binding properties of both exogenously expressed and recombinant RNase H1 toward R-loop. According to this strategy, a recombinant protein complex, named RNA-DNA hybrid biosensor, consisting of two tandom HBDs (2xHBD) fused to GST (glutathione S-transefease) and hexahistidine ($His_6$) tags was used for capturing R-loops **(K. Wang *et al.*, 2021)**. This strategy was used to map R-loops with two different methods; in the first, genomic material was extracted and enzymatically digested then purified and incubated with the biosensor complex for enrichment befor pull down using glutathione beads. After DNase digestion, the

final library was prepared from RNA similar to S9.6-DRIPc. Although this method enriched *in vitro* R-loops (similar to DRIVE-seq which has some disadvantages), the generated profiles were reported to show concordance with indepedent S9.6-DRIPc profiles. Beside delivering stranded signals, this method was reported to have high resolution and also high sensitivity compared to DRIPc which uses RNA template for sequencing **(K. Wang *et al.*, 2021)**.

To map *in situ* R-loops using native conditions and to overcome the potential bias of enzymatic digestion, the second method used the CUT&Tag (Tn5-based cleavage under targets and tagmentation) **(Kaya-Okur *et al.*, 2020; Kaya-Okur *et al.*, 2019)** approach. This approach incorporates the use of Tn5 transposase that can cleave the DNA template and directly ligate sequencing-compatible adaptors to it. For R-loop mapping with CUT&Tag method, bead-immobilized cells were permeabilized before sequential incubations with the biosensore complex, compatible antibodies and finally recombinat proteinA-Tn5 protein complex. The authors also tried the use of CUT&Tag method in combination with S9.6 instead of the HBD biosensor complex. They reported that both combinations of the CUT&Tag method (with either S9.6 or HBD) generated highly similar native R-loop profiles **(K. Wang *et al.*, 2021)**.

An advantage for the CUT&Tag method is the use of native *in vivo* (*in situ*) conditions to map R-loops. Also, the use of tagmentation was reported to increase the sensitivity and, enhance the resolution compared to DRIP-ChIP and other RNase H-based methods which use DNA template for sequencing. Remarkably, distinct R-loop profiles were found to be generated with the the same R-loop sensor (S9.6 or RNase H) depending on whether *in situ* or *ex vivo* conditions were used, which led to the conclusion that discrepancies revealed by different R-loop mapping methods are attributed to differences in capturing conditions rather than specificity of sensors **(K. Wang *et al.*, 2021)**.

Finally, some limitions have been reported against the use of Tn5 transposase. It has been reported to have high affinity toward accessible DNA which increases the background necessitating the need for stringent washing that may underrepresent some signals. Also, the pA-Tn5 is much bulkier compared to pA-MNase, for example, which may present steric hinderance and impact resolution **(Kaya-Okur *et al.*, 2020)**. Interestingly, Tn5 has been reported to have direct tagmentation activity on RNA-DNA hybrids as it can ligate adaptors directly to both of the DNA and RNA strands of R-loops **(Di *et al.*, 2020; Lu *et al.*, 2020)**.

# Chapter 2. Materials and Methods

## 2.1 General molecular biology techniques and fission yeast methods

### 2.1.1 Standard PCRs and restriction digestion

Unless otherwise stated, all standard PCRs were carried out using Promega GoTaq® DNA Polymerase (M3001) in 20 µl reaction volume following recommended manufacturer program. Unless stated otherwise, primers were manufactured by Sigma Aldrich. Restriction digestions were performed using NEB enzymes following NEB protocols in 50 µl reaction volume, or in lower volumes with scaling reaction components volumes. PCR amplification or restriction digestion products were loaded on 1% agarose gel pre-stained with either 1X red safe (Bioline) or 1X stain G (Serva) and run till complete separation of bands before visualization. Unless stated otherwise, 1 µg of 2-log ladder (NEB) was used for PCR band and fragment size determination.

### 2.1.2 Quantitative PCRs (qPCRs)

Before quantitative real time PCRs (qPCRs), DNA was serially diluted, and sets of triplicates were used for each condition. qPCR reactions were prepared using NEB Luna® Universal qPCR Master Mix (M3003S) in 20 µl reaction volume containing 10 µl of 2X master mix, 250 nM of each primer and 5 µl of diluted DNA. Reaction plates were run on 7900HT (Applied Biosciences) real time instrument with program set for initial denaturing at 95 °C for 1 min, denaturing at 95 °C for 15 sec and extension at 60 °C for 30 °C. Primers used for qPCRs are listed in Table 2.1. Melting curve was generated using instrument-recommended conditions with 60 °C for annealing and 95 °C for denaturing.

**Table 2.1. List of primers used for qPCR experiments.**

| No. | Name | Sequence | Target sequence |
|---|---|---|---|
| A280 | qPCR.kre33.fw | AGTTAAGGCAGAGCGACAGA | 5’ ETS of yeast 35S rDNA |
| A281 | qPCR.kre33.rv | TCCCAACTACTTTTCCTCACAC | 5’ ETS of yeast 35S rDNA |

For checking differential enrichment for PCR signals on agarose gel, target DNA was amplified following standard PCR conditions above and using the same target specific primers and primers concentrations for qPCR reactions. Minimal cycling (28 to 30 cycles) was applied to differentiate signals strengths after agarose gel electrophoresis.

### 2.1.3 Yeast cell lysis and cell extract preparation

*S. pombe* cells used for R-loop project-related work are listed in Table S1. Equal number of yeast cells, equivalent to 10 ODs at $OD_{600}$, were used for whole cell extract preparation. Cells were collected from freshly growing liquid cultures and centrifuged at 4500 rpm for 2 min at room temperature then resuspended in 1 ml water and moved to a screw-cap tube. After centrifuging at 6000 rpm and aspirating the supernatant, cell pellet was either snap frozen in liquid nitrogen or kept on ice to be processed directly. To the pellet was added 200 µl of cold 0.5 mm zirconium beads (Biospec) and 400 µl cold 20% Trichloroacetic acid (TCA) before homogenising in Precellys® 24 tissue homogenizer (Bertin) for 4 times, each for 20 sec at 5000 rpm with 5-sec pause in between the cycles. The cell lysate was centrifuged at 13000 rpm for 5 min at 4 °C. After carefully removing the supernatant, the beads with the pellet were dried using a speed vacuum at 30 °C for 30 – 50 min, or allowed to dry overnight. 150 µl of 1 M Tris pH 9.4 and 150 µl of 4X lämmli buffer (250 mM Tris pH 6.8, 9.2% SDS, 40% Glycerol and 0.2% bromophenol blue) including 0.1 M DTT were added to the beads/pellet at room temperature before vortexing for 1 min then heating at 95 °C for 5 min. Finally, the samples were centrifuged at 14000 rpm for 5 min and the supernatant was moved to a new tube to be loaded on SDS PAGE gel.

### 2.1.4 SDS-PAGE

For SDS PAGE gel electrophoresis, Invitrogen Mini Gel tanks and Novex™, NuPAGE™ 4-12% Bis-Tris Mini Protein Gels were used. Just before the run, the samples were heated at 37 - 42 °C for 2- 3 min during which the gel tank was filled with 1X MOPs buffer and 500 µl antioxidant. After assembling the components, equal volumes of samples were loaded on gel, and empty lanes were filled with the same volume of 1X Lämmli buffer. Equal volume of Color Prestained Protein Standard, Broad Range (11–245 KD) (NEB #P7719) ladder was loaded to verify bands molecular weight. Protein separation was done at 100 V for 5 min then at 200 V till enough separation.

### 2.1.5 Western blot

For semi-dry Western blotting, iBlot™ device and iBlot™ 2 mini transfer stacks with nitrocellulose membrane were used. The gel and filter paper were assembled along with the transfer stack according to manufacturer manual, and the transfer step was done using program

P3 for 7-8 min. For wet Western blotting, nitrocellulose membrane, sponge pads and filter papers were immersed separately in 1X transfer buffer (25 mM Tris, 192 mM Glycerine, 20% Methanol) before stacking with the gel using Invitrogen Xcell II transfer module. The transfer module was lodged in the mini tanks and the inside of the module was filled with 1X transfer buffer while the tank was filled with distilled water. The transfer was carried out at 30 V for 2 hours.

At the end of the transfer, the membrane was carefully removed and shortly stained with Ponceau red solution then washed with distilled water. For blocking, the membrane was immersed with 5% milk in 1X PBST (1X PBS, 0.1% Tween-20) solution for 1 hour with shaking on a bench swing shaker. The milk was poured to stop the blocking, and the membrane was incubated with the primary antibody diluted in 5% milk-1X PBST for one hour at room temperature or overnight at 4 °C. All primary antibodies (listed in Table S2) were used at 1:1000 dilution except anti-hexokinase antibody which was used at 1:10,000 dilution. Finally, the membrane was washed twice in PBST then once in 1X PBS, each for 15 min with gentle shaking before incubating with diluted secondary antibody for one hour at room temperature. After repeating washing steps, Immobilon Western Chemiluminescent HRP Substrate (Millipore) was added to the membrane for signal detection and visualization with Amersham™ imager.

## 2.2 Immunoprecipitation, affinity pulldown and library preparation methods

All solutions and buffers used for yeast or mammalian cell lysis, chromatin sonication and chromatin immunoprecipitation or affinity pulldown experiments are listed in Table 2.2. Autoclaved MQ (Millepore) H2O was used for preparing all solutions and buffers. Fresh and cold work buffers were used.

### 2.2.1 Yeast cells crosslinking, lysis and chromatin sonication (chromatin preparation)

*S. pombe* or *S. cerevisiae* yeast cells were grown overnight in 250 ml of media to $OD_{600}$ of 0.8 - 1.0 (mid-log growth phase) then crosslinked directly in the same culture using fresh formaldehyde (37% FA-15% MeOH, Sigma Aldrich) at 1% final FA concentration by shaking at 150 rpm for 15 min at room temperature. ***Crosslinking was skipped for non-crosslinked samples.*** Cells were pelleted by centrifugation at 2800 xg for 3 min at 4 °C then washed twice using enough ice cold 1X PBS with centrifugation at 2800 xg for 3 min at 4 °C in between.

Finally, the pellet was suspended and moved to a screw-cap cryotube to be centrifuged at 2300 xg for 3 min at 4 °C. Supernatant was discarded and cells were snap-frozen in liquid Nitrogen and kept at -80 °C to the time of use.

For all the next steps, samples were kept on ice whenever possible, centrifugation was done at 4 °C and fresh ice-cold buffers were used unless otherwise stated. Cells were thawed on ice, resuspended in 750 µl ice-cold FA-SDS lysis buffer, split into two halves and moved to two screw cap tubes filled with cold 500 µl zirconium beads (Biospec) for bead beating. Lysis was done with Precellys® 24 tissue homogenizer (Bertin) using parameters; three times, 20 sec and 5500 rpm in a cold room. Samples were kept on ice for 2 min between each beating cycle. Tubes containing homogenized samples were pierced using hot needles, inserted in 5-ml glass tubes and centrifuged at 200 xg for 1 min at 4 °C to bring cell lysate down.

After resuspension, cell lysates from two similar tubes (split earlier during lysis) were combined together then moved to 1-ml fibre-cap Covaries tubes and kept on ice till sonication ends. Water bath of Covaries device was filled with 1600 ml water to be cooled and degassed before sonication. Sonication was done using Covaries S2 machine following recommended operating conditions and using parameters; cycles per burst (CBP) 200; intensity (I) 5; and duty factor (DF) 20% for 8 min on frequency sweeping mode to generate peak power of 175 and DNA average fragment size of 200 – 300 bp. Sonicated samples were moved to 1.5-ml Eppendorf tubes and centrifuged at 2300 xg for 10 min at 4 °C to pellet cell debris. Sonicated chromatin was moved to fresh Eppendorf tube and diluted with equal volume of FA lysis buffer plus protease inhibitors (PIs) without SDS.

### 2.2.2 Mammalian cells crosslinking, nuclei isolation and chromatin sonication

After reaching desired confluency, culture media were poured, and cells were fixed in 1% FA-1X PBS buffer by rocking on a swinging shaker for 10 min at room temperature. Fixation was stopped by adding 0.125 M glycine then rocking the same way for 5 min. Fixation buffer was disposed, and cells were washed twice with 10 ml ice cold 1X PBS each. After discarding PBS, cells were scrapped and moved to 50-ml falcons using 1X PBS buffer (including fresh PIs). Cells were pelleted by centrifugation at 2500 xg for 5 min before aspirating liquid. Cell pellet was resuspended in 1X PBS and moved to 2-ml screw cap tube for centrifugation at 2500 xg for 5 min before snap freezing in liquid nitrogen. Lysis was done using 2 ml of mammalian cell lysis buffer and shaking on a rotating wheel at 4 C for 15 min.

Nuclei were pelleted by centrifugation at 2500 xg for 5 min before aspirating liquid then washed using 1.5 ml nuclei wash buffer by resuspending using wide-pore tip and rotation for 5 min at 4 °C. After pelleting again by centrifugation at 2500 xg for 5 min, nuclei were lysed using 1 ml sonication buffer, resuspended and rotated for 5 min at 4 °C. Tubes were spinned shortly and transferred to 1-ml AFA fiber cap Covaries tube for sonication. Sonication was done using Covaries S2 device following parameters; CPB 200; I 5; and DF 20% for 9 min. Sonicated chromatin was moved to 1.5-ml Eppendorf tube then 50 µl of 20% triton X-100 (1% final) was added and mixed well before centrifuging at 16000 xg for 10 min at 4 °C. Chromatin was diluted 1:2 using IP buffer 1 before incubating with required antibody.

### 2.2.3 Mammalian cells whole cell extract preparation and sonication

Low-input mammalian cells samples were crosslinked, quenched and washed as mentioned above for mammalian cells. For lysis, cells were resuspended in 140 µl lysis & sonication buffer and kept on ice for 20 min with intermittent hand shaking every 5 min. Lysed cells were moved to 130-µl FA fiber slit-cap Covaries tubes for sonication following the above-mentioned parameters for mammalian cells. After sonication, chromatin was moved to a fresh Eppendorf tube to be centrifuged at 16000 xg for 10 min at 4 °C. Chromatin was diluted 1:10 with IP buffer 2 before incubating with required antibody.

### 2.2.4 Chromatin immunoprecipitation

Immunoprecipitation was done by incubating sonicated chromatin overnight with 5 to 10 µg of antibody at 4 °C on a rotating wheel. A no-antibody control sample was included for each experiment. Prior to affinity pulldown for antibody-bound chromatin, a protein-A/protein-G sepharose-bead mix (sepharose fast flow, Sigma) was prepared in a batch for all samples at a 1:1 ratio by calculating 40 µl total volume of bead slurry for each sample (20 µl Protein-A + 20 µl Protein-G). Cut-end tips were used in all steps involving handling beads. Bead mix was washed three times, each with 1 ml FA lysis buffer plus PIs without SDS by gentle hand inversion followed by centrifugation at 94 xg for 1 min at 4 °C. After the final centrifugation, buffer was removed and beads were diluted with a fresh buffer. Equal volume of beads equivalent to 40 µl slurry was added to each chromatin-antibody sample and incubated for 2 hours at 4 °C on a rotating wheel.

At the end of bead incubation, samples were centrifuged at 94 xg for 1 min at 4 °C to bring the beads down, and supernatant was carefully aspirated without touching the beads. Beads were gently resuspended in fresh FA lysis buffer plus PIs without SDS, using a cut-end tip, moved to a fresh Eppendorf tube and centrifuged the same way. Beads were washed twice, each with 1 ml of FA lysis buffer plus PIs without SDS, by incubating on a turning wheel for 5 min followed by centrifuging at 94 xg for 1 min at 4 °C. Washing was repeated the same way for one time using 1 ml of each of the following buffers in order; ice-cold wash buffer 1; ice-cold wash buffer 2; ice-cold wash buffer 3; and 1X TE buffer. Samples were centrifuged at 68 xg for 1 min at room temperature and supernatant was removed in between before adding the next buffer. After washing, different reactions were performed on beads based on different ChIP-exonuclease experiments. These reactions are to be detailed in the next subsections.

### 2.2.5 SDS elution

At the end of enzymatic reactions and before elution, beads were washed only in 1X TE buffer without the 10 mM Tris rinse step. Elution was done by adding 100 µl of SDS elution buffer to the beads and mixing by pipetting then incubating in a thermomixer at 65 °C with shaking at 1000 rpm for 15 min. Beads were centrifuged at 1200 xg for 1 min at room temperature. Eluate was aspirated carefully without touching the beads and moved to a fresh tube. Elution was repeated, and eluates of one sample were combined before adding 250 µl of 1X TE buffer. RNA was eliminated by adding 2 µl of RNase A/T cocktail (Ambion) and incubating at 37 °C for 1 hour. Decrosslinking was done by adding 2 µl of proteinase K (Ambion) and incubating overnight in a 65 °C incubator.

### 2.2.6 Phenol chloroform extraction for DNA

Samples were allowed to cool down to room temperature before adding equal volume of 25:24:1 phenol/chloroform/isoamyl (Sigma) for DNA extraction, mixing vigorously for 1 min and centrifuging at 18000 xg for 6 min. The upper liquid phase was moved to a fresh tube before adding to it an equal volume of chloroform and repeating the previous step. Prior to DNA precipitation, 40 mg glycogen (Serva), 200 mM NaCl and 1 ml 100% EtOH (-80 °C) were added to the eluate, mixed quickly and incubated at (-80 °C) for 2 hours then centrifuged at maximum speed for 1 hour at 4°C. liquid was removed and pellet was washed with 500 ml cold 75% EtOH before centrifuging at maximum speed for 20 min at 4°C. liquid was removed

again and the pellet was allowed to dry at room temperature before dissolving in 11 µl TE and moving to PCR tubes.

**Table2.2. Solutions and buffers used for cell lysis, sonication and immunoprecipitation.**

| Buffer | Components | Usage |
|---|---|---|
| **FA-SDS & FA lysis buffers** | 50 mM HEPES-KOH pH 8.0, 2 mM EDTA pH 8.0, 150 mM NaCl, 1% triton X-100, 0.1% sodium deoxycholate (10 mg), 1X FY protease inhibitor, 1 mM PMSF, 0.1% SDS (FA-SDS buffer only). | Yeast cells lysis and sonication |
| **FA wash buffer 1** | 50 mM HEPES-KOH pH 8.0, 2 mM EDTA pH 8.0, 1M NaCl, 1% triton X-100, 0.1% sodium deoxycholate (10 mg). | Common buffers for yeast and mammalian cells ChIP |
| **FA wash buffer 2** | 50 mM HEPES-KOH pH 8.0, 2 mM EDTA pH 8.0, 500 mM NaCl, 1% triton X-100, 0.1% sodium deoxycholate (10 mg). | |
| **FA wash buffer 3** | 10 mM Tris-Cl pH 8.0, 2 mM EDTA pH 8.0, 25 mM LiCl, 0.1% Nonidet P-40 (NP-40), 1% sodium deoxycholate (100 mg). | |
| **General wash buffer** | 10 mM Tris-Cl pH 8.0, 2 mM EDTA pH 8.0, 300 mM NaCl, 0.1% triton X-100. | |
| **SDS elution buffer** | 25 mM Tris, pH 8.0, 10 mM EDTA pH 8.0, 200 mM NaCl, 1% (w/v) SDS. | |
| **TEV elution buffer** | 50 mM Tris pH 8.0, 0.5 mM EDTA pH 8.0, 100 mM NaCl, 10% glycerol, 0.1% Igepal, 0.5 mM DTT. | |
| **3X FLAG peptide elution** | 50 mM Tris pH 7.5,10 mM EDTA pH 8.0, 200 mM NaCl. | |
| **Mammalian cell lysis buffer** | 50 mM Hepes pH 7.5, 1 mM EDTA, 140 mM NaCl, 10% glycerol, 0.5% NP-40, 0.25% triton X-100, 1mM PMSF, 1X Roch protease inhibitor. | Mammalian cells |
| **Nuclei wash buffer** | 10 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl, 1 mM PMSF, 1X Roch protease inhibitor. | |
| **Sonication buffer** | 10 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, 0.1% sodium deoxycholate, 1 mM PMSF, 0.5% sarcosine, 1X Roch protease inhibitor. | |
| **IP buffer 1** | 10 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, 0.1% sodium deoxycholate, 1 mM PMSF, 1% triton X-100, 1X Roch protease inhibitor. | |
| **Lysis & sonication** | 10 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS, 1 mM PMSF, 1X Roch protease inhibitor. | Low-input mammalian cell samples |
| **IP buffer 2** | 16.7 mM Tris-HCl pH 8, 1.2 mM EDTA, 167 mM NaCl, 0.11% sodium deoxycholate, 1 mM PMSF, 0.01% SDS, 1.1% triton X-100 1X Roch protease inhibitor. | |

### 2.2.7 Chromatin affinity pulldown for FTP-tagged proteins

Yeast cells carrying FTP-tagged proteins were fixed, lysed and sonicated as explained above. Immunoaffinity pulldown was carried out by incubating chromatin with 40 µl IgG beads slurry (sepharose 6 fast flow, Sigma) for 2 hours at 4 C° with rotation. Before adding to chromatin, beads were washed in a batch three times, each with 1 ml of FA-lysis buffer plus PIs without SDS, then diluted before adding equal volume to each sample. At the end of bead incubation, samples were centrifuged at 94 xg for 1 min at 4 °C to bring the beads down, and supernatant was carefully aspirated without touching the beads. Beads were gently resuspended in fresh FA lysis buffer plus PIs without SDS using a cut-end tip, moved to a fresh Eppendorf tube and centrifuged the same way. Beads were washed once with 1 ml of FA lysis buffer plus PIs without SDS by incubating on a turning wheel for 5 min followed by centrifugation at 94 xg for 1 min at 4 °C.

### 2.2.8 TEV elution and FLAG pulldown

After washing once with 1 ml 1X TE buffer, beads were centrifuged at 68 xg for 1 min at room temperature, and supernatant was removed before the TEV elution. TEV elution was performed by adding 200 µl TEV elution buffer and 5 µl self-made TEV protease to the beads then incubating at 4 ºC overnight with rotation. TEV eluate was moved to a fresh tube after centrifuging at 135 xg for 1 min. 50 µl anti-FLAG M2 (Sigma) affinity bead slurry was added to TEV eluate after washing and diluting the same way above. 1 ml of FA lysis buffer was added to TEV eluate-bead mix before incubating for 3 hours at 4 °C with rotation. Beads were washed sequentially using washing buffers as indicated above for chromatin immunoprecipitation. After washing, different reactions were performed on beads toward FLAG elution.

### 2.2.9 FLAG elution

FLAG elution was done by adding 150 µl of 3X FLAG Peptide elution buffer and 10 µl 3X FLAG Peptide (Sigma) to the beads and mixing by inverting then incubating on a turning wheel at 4 ºC for 30 min. Beads were centrifuged at 1200 xg for 1 min at room temperature. Eluate was aspirated carefully without touching the beads and moved to a fresh tube. Elution was repeated and eluates of one sample were combined before adding 250 µl of 1X TE buffer. RNA was eliminated by adding 2 µl of RNase A/T cocktail (Ambion) and incubating at 37 °C

for 1 hour. Decrosslinking was done by adding 2 µl of proteinase K (Ambion) and incubating overnight at 65 °C. Phenol chloroform extraction for DNA was done as mentioned above.

### 2.2.10 Original ChIP-exonuclease (ChIP-exo 1)

Original ChIP-exonuclease (also called ChIP-exo 1) was performed according to a previously published protocol **(Rhee & Pugh, 2012)** with some modifications. The ChIP-exo protocol includes the chromatin immunoprecipitation steps (previous sections) plus the enzymatic reactions toward the library preparation. Part of these enzymatic reactions is carried out on sepharose beads (on-bead) before elution while the other part is carried out after elution and DNA precipitation (in-solution). All enzymes and buffers for enzymatic reactions without exception were purchased from NEB.

#### *2.2.10.1 On-bead enzymatic reactions*

Unless otherwise stated, the following conditions were adopted from now on for all the ChIP-exo versions; at the end of the washing cycles, tubes containing beads were always kept on ice with beads submerged in enough volume of 1X TE till the next step; prior to every enzymatic reaction, beads were equilibrated by rinsing (washing without inversion) with a 10 mM Tris-HCl solution of pH similar to the next reaction buffer; tubes were centrifuged at 68 xg for 1 min at room temperature, and all supernatant was removed; ~40 µl of the reaction mix was added on the surface of the beads without pipetting or dispersing the beads on the wall of the tubes; reaction mix was prepared for all the samples for every reaction for a total reaction volume of 60 µl per sample; at the end of each reaction, samples were rinsed with 1 ml general wash buffer, rinsed twice with the same buffer and once with 1X TE buffer; each washing cycle was done for 5 min with rotation on a turning wheel at 4 °C; and all centrifugations were done at 68 xg for 1 min at room temperature.

Reactions components for ChIP-exo 1 are listed in Table 2.3. For all samples, beads were incubated sequentially in the following reactions for the specified times and temperatures; DNA-end polishing (blunting) at 12 °C for 20 min with shaking in a thermomixer at 500 rpm; first kinase (phosphorylation) at 37 °C for 30 min; 3'end dA-tailing at 37 °C for 30 min; first adaptor ligation at 16 °C overnight followed by; second kinase reaction at 37 °C for 30 min prior to; lambda exonuclease reaction to digest 5' ends of dsDNA, which was done at 37 °C for 30 min for the crosslinked or for 6 min for the non-crosslinked samples; and, finally, $Rec_{Jf}$

reaction to digest 5' ends of ssDNA, which was done at 37 °C for 30 min for the crosslinked samples. The $Rec_{Jf}$ reaction was skipped for the non-crosslinked samples. Washings and centrifugations were done after all the steps as specified above except for the first adaptor ligation where beads were sequentially washed; twice using FA-lysis buffer without PIs without SDS; and once with each of the FA washing buffers.

#### *2.2.10.2 RNase H treatment for R-loop mapping experiments*

For R-loop mapping experiments, control sample was treated with RNase H (RNase H treatment control) before the above reactions. Beads were rinsed with 1 ml 10 mM Tris-HCl pH 8, tubes were spinned as indicated, and all supernatant was aspirated carefully. Bead-bound chromatin was incubated in RNase H reaction (1X RNase H buffer and 20 U RNase H) (NEB) for 30 min at 37 °C on a rotating wheel. At the end of incubation, samples were brought to room temperature before putting on ice. Beads were rinsed once with 1 ml general wash buffer then washed twice with the same buffer before washing with 1X TE buffer.

#### *2.2.10.3 In-solution enzymatic reactions*

After phenol chloroform extraction and DNA precipitation, extension primer was annealed by incubating DNA with denaturing-annealing mix in a PCR machine using program; 95 °C / 5 min (denaturing) - 62 °C / 5 min (annealing) - cooling down to room temperature. 1 µl of 10 U/µl phi29 DNA polymerase was added in a final 20 µl reaction volume, and extension was done in a PCR machine using program; 30°C / 20 min (primer extension) - 65°C / 10 min (heat inactivation) - 4 °C. For the second dA-tailing, 10 µl dA-tailing mix was added to the DNA extension product (30 µl total) and incubated in a PCR machine at 72 °C for 20 min then allowed to cool down to room temperature. DNA was moved to an Eppendorf tube for AMPure XP beads purification. Bead purification was done according to manufacturer manual using 42 µl AMPure XP beads (1.4 volume of DNA). DNA was dissolved in 30 µl TE buffer and moved to PCR tubes for the second adaptor ligation which was done in 40 µl reaction volume in a PCR machine at 16 °C overnight.

**Table 2.3. List and components of enzymatic reactions conducted for ChIP-exo 1 (two A-T ligations).** Volumes are inserted in parentheses and final concentrations in straight brackets. All reactions volumes are in (µl) unless stated otherwise. All on-bead reactions were done in 60 µl volume while *Taq* and Phusion PCR reactions were done in 20 and 25 µl volumes, respectively.

| No. | Reaction | Components |
|---|---|---|
| **1** | **Polishing/blunting** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 8.0 (27 µl), 10X NEB buffer 2 (6 µl) [1X], 10X BSA (3 µl) [0.5X], 3 mM dNTPs (3 µl) [150 µM each], 3 U/µl T4 DNA polymerase (1 µl) [3 U]. |
| **2** | **First kinase** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 7.5 (33 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 10 U/µl T4 PolyNucleotide Kinase (PNK) (1 µl) [10 U]. |
| **1 & 2** | **Polishing & kinase for improved ChIP-exo 1** | Beads with chromatin (20 µl), H2O (20.5 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 3 mM dNTPs (8 µl) [400 µM], 3 U/µl T4 DNA polymerase (2.5 µl) [7.5 U], 5 U/µl Large Klenow (0.5 µl) [2.5 U], 10 U/µl T4 PNK (2.5 µl) [25 U]. |
| **3** | **First dA-tailing** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 8 (31 µl), 10X NEB buffer 2 (6 µl) [1X], 3 mM dATPs (2 µl) [100 µM], 5 U/µl Klenow exo (1 µl) [5 U]. |
| **4** | **First A-T adaptor ligation** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 7.5 (28 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 15 µM Uni-DS short adaptor (5 µl) [1.25 µM], 400 U/µl T4 DNA ligase (1.25 µl) [500 U]. |
| **5** | **Second Kinase** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 7.5 (33 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 10 U/µl T4 PNK (1 µl) [10 U]. |
| **6** | **Lambda exonuclease** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 9.2 (32 µl), 10X λ exo buffer (6 µl) [1X], 5U/µl λ exonuclease (3 µl) [15 U]. |
| **7** | **Rec$_{Jf}$ exonuclease** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 8 (33 µl), 10X NEB buffer 2 (6 µl) [1X], 30 U/µl Rec$_{Jf}$ exonuclease (1.5 µl) [45 U]. |
| **Elution, RNA digestion, decrosslinking, and DNA extraction and precipitation** | | |
| **8** | **Denaturing and annealing** | DNA (11 µl), 10X Phi 29 DNA Pol. buffer (2 µl) [1X], 10X BSA (4 µl) [2X], 3 mM dNTPs (1 µl) [150 µM], 20 µM Uni-ext extension primer (1 µl) [1 µM]. |
| **9** | **Primer extension** | DNA (11 µl), 10X Phi 29 DNA Pol. buffer (2 µl) [1X], 10X BSA (4 µl) [2x], 3 mM dNTPs (1 µl) [150 µM], 20 µM Uni-ext extension primer (1 µl) [1 µM], 10 U/µl Phi 29 DNA Pol. (1µl) [10 U]. |
| **10** | **Second dA-tailing** | DNA (20 µl), 1X TE (4.5 µl), 10X *Taq* buffer (3 µl) [1X], 3 mM dATPs (2 µl) [200 µM], 5 U/µl *Taq* Pol. (0.5 µl) [2.5 U]. |
| **11** | **Second A-T ligation** | DNA (30 µl), 1X TE (4 µl), 10X T4 DNA ligase buffer (4 µl) [1X], 15 µM NL5Bc-DS adaptor (1 µl) [0.375 µM], 400 U/µl T4 DNA ligase (1.25 µl) [500 U]. |
| **12** | ***Taq* PCR check** | H2O (12.5 µl), 5X Promega *Taq* buffer (4 µl), 10 mM dNTPs (0.4 µl) [200 µM], 20 µM PCR-fw primer (0.5 µl) [0.5 µM], 20 µM PCR-rv primer (0.5 µl) [0.5 µM], Promega *Taq* DNA polymerase (0.25 µl) [1.25 U], 2 µl DNA. |

### 2.2.11 Improved ChIP-exo 1

Improved ChIP-exo 1 was done identical to the above protocol with the exception that the first polishing and kinase steps were combined in one step (Table 2.3) and samples were incubated at 20°C for 30 min with rotation. This step was found to increase the efficiency of the whole experiment, and hence used in all ChIP-exo experiments with two A-T ligation steps.

### 2.2.12 ChIP-exo-(SL)2 enzymatic reactions (two splint ligations)

ChIP-exo-(SL)2 was performed according to a recent publication **(Rossi *et al.*, 2018)**, where it's coined as ChIP-exo 4, with some modifications. Best working conditions and reactions components of ChIP-exo-(SL)2 are listed in Table2.4. ***For on-bead reactions***, polishing and kinase reactions (separate or combined) were done identical to original ChIP-exo 1 protocol detailed above. Directly, after this reaction, lambda exonuclease and Rec$_{Jf}$ reactions were carried out sequentially, each at 37 °C for 30 min on a turning wheel. After these reactions, splint ligation was done at 16 °C overnight to ligate the long barcoded NL5Aa-SL splint adaptor to the 3' end of DNA. Beads were stringently washed before SDS elution, RNase A/T treatment, decrosslinking, and DNA extraction and precipitation which were done identical to original protocol but with dissolving DNA pellet in 20 µl 1X TE buffer. ***In-solution reactions*** involve only one splint ligation step which was performed at 16 °C overnight to ligate the short Uni-SL splint adaptor to the 5' end of DNA. Prior to ligation, the DNA was denatured at 95 °C for 3 min then quickly chilled on ice. The experiment was continued to the end similar to original ChIP-exo 1 protocol.

### 2.2.13 ChIP-exo-SL-DS enzymatic reactions (splint then A-T ligation)

ChIP-exo-SL-DS is a combination of ChIP-exo-(SL)2 and ChIP-exo 1. On-bead reactions toward first splint ligation were done similar to ChIP-exo-(SL)2 while in-solution reactions toward A-T ligation were done similar to ChIP-exo 1 (Table S3).

### 2.2.14 ChIP-exo-DS-SL enzymatic reactions (A-T then splint ligation)

ChIP-exo-DS-SL is a combination of ChIP-exo 1 and ChIP-exo-(SL)2. On-bead reactions toward first A-T ligation were done similar to original ChIP-exo 1 while in-solution reactions toward splint ligation were done similar to ChIP-exo-(SL)2 (Table S4).

**Table 2.4. List and components of enzymatic reactions conducted for ChIP-exo-(SL)2 (two splint ligations).**

| No | Reaction | Components |
|---|---|---|
| **1** | **Polishing/blunting** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 8.0 (27 µl), 10X NEB buffer 2 (6 µl) [1X], 10X BSA (3 µl) [0.5X], 3 mM dNTPs (3 µl) [150 µM each], 3 U/µl T4 DNA polymerase (1 µl) [3 U]. |
| **2** | **First kinase** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 7.5 (33 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 10 U/µl T4 polynucleotide kinase (1 µl) [10 U]. |
| **3** | **Lambda exonuclease** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 9.2 (32 µl), 10X λ exo buffer (6 µl) [1X], 5 U/µl λ exonuclease (4 µl) [20 U]. |
| **4** | **Rec*Jf* exonuclease** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 8 (33 µl), 10X NEB buffer 2 (6 µl) [1X], 30 U/µl Rec*Jf* exonuclease (2.5 µl) [75 U]. |
| **5** | **First adaptor ligation** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 7.5 (28 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 5 µM Nl5Aa-SL splint adaptor (4.5 µl) [0.375 µM], 400 U/µl T4 DNA ligase (3 µl) [1200 U]. |
| **Elution, RNA digestion, decrosslinking, and DNA extraction and precipitation** | | |
| **6** | **Second splint ligation** | DNA (30 µl), 1X TE (4 µl), 10X T4 DNA ligase buffer (4 µl) [1X], 10 µM Uni-SL (1.5 µl) [0.375 µM], 400 U/µl T4 DNA ligase (3 µl) [1200 U]. |

### 2.2.15 ChIP-exo-TT enzymatic reactions (poly dA-tailing and A-T ligation)

ChIP-exo terminal transferase (ChIP-exo-TT) enzymatic reactions and their components are listed in Table 2.5. ***On-bead reactions*** included only the combined blunting & kinase treatment followed by the lambda exonuclease treatment. In some experiments, both the blunting and kinase were omitted, and the lambda exonuclease was replaced with a T7 exonuclease treatment at 37 °C for 30 to 45 min before elution. After DNA precipitation, DNA pellet was dissolved in 18 µl H2O and moved to a PCR tube. ***In-solution reactions*** started with a poly-dA tailing step which was done in a thermocycler using program; 37 °C / 30 min - 70 °C / 10 min - 20 °C. DNA was purified with AMPure XP beads of 1.8 DNA volume. Extension was conducted as above but using Uni-dT24VN primer. A control sample without terminal deoxy-transferase enzyme (no-TdT) was included. Reactions were conducted toward double stranded A-T ligation similar to original ChIP-exo 1.

**Table 2.5. List and components of enzymatic reactions conducted for ChIP-exo-TT.**

| No | Reaction | Components |
|---|---|---|
| **1** | **Combined polishing & kinase** | Beads with chromatin (20 µl), H2O (20.5 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 3 mM dNTPs (8 µl) [400 µM], 3 U/µl T4 DNA polymerase (2.5 µl) [7.5 U], 5 U/µl Large Klenow (0.5 µl) [2.5 U], 10 U/µl T4 PNK (2.5 µl) [25 U] |
| **2** | **Lambda exonuclease** | Beads with chromatin (20 µl), 10 mM Tris-Cl, pH 9.2 (32 µl), 10X λ exo buffer (6 µl) [1X], 5U/µl λ exonuclease (3 µl) [15 U] |
| **3** | **T7 exonuclease** | Beads with chromatin (20 µl), H2O (32 µl), 10X NEB buffer 4 (6 µl) [1X], 10 U/µl T7 exonuclease (2 µl) [20 U] |
| **Elution, RNA digestion, decrosslinking, and DNA extraction and precipitation** | | |
| **4** | **Poly-A tailing** | DNA (18.5 µl), 10X TdT buffer (2.5 µl) [1X], 10X CoCl2 (2.5 µl) [1X], 5 mM dATPs (0.5 µl) [100 µM], 20 U/µl NEB TdT (1 µl) [20 U], |
| **Annealing, extension, dA tailing and second ligation reactions are similar to ChIP-exo 1** | | |

### 2.2.16 ChIP-exo-SMART and low-input ChIP-exo-SMART

On-bead reactions for ChIP-exo-SMART included only the T7 exonuclease reaction. After elution and DNA precipitation, poly-dA tailing was conducted as in ChIP-exo-TT. Poly-dA tailing was followed by AMPure XP beads purification for DNA which was dissolved in 15 µl H2O and moved to a PCR tube. To the DNA was added 1 µl of 10 mM dNTPs and 1 µl of 2 µM Uni-dT24VN extension primer (17 µl final). Denaturing of DNA and primer annealing was done by running program 95 °C / 3 min – 55 °C / 3 min then quickly cooling on ice. A master mix of; 5 µl of 5X first strand buffer, 2 µl of 100 mM DTT and 1 µl of 200 U/µl SuperScript™ II (8 µl total) was added to the DNA/primer mix in a final 25 µl reaction volume. Reaction mix was incubated at 42 °C for 15 – 20 min then brought to room temperature before adding 1 µl of 10 µM TSP primer. Reaction components were incubated again at 42 °C for an hour before denaturing at 70 °C for 15 min.

***For low-input samples***, before the dA-tailing, DNA was treated with 1 U of rSAP in 1X TdT buffer in 15 µl reaction volume using program; 37 °C / 15 – 20 min - 65 °C / 5 min - 4 °C, directly, before adding the rest of the poly-dA reaction components. After the tailing reaction, and before the reverse transcription, DNA was purified using AMPure XP beads of 1.8 DNA volume. All oligos and primers used for the different ChIP-exo experiments are listed in Table 2.6.

**Table 2.6. List of oligos and primers used for ChIP-exo and SMART methods.** Different end modifications are included (blue) when applicable. In-line barcodes (Bc) of variable lengths (green) and 6Ns unique molecular identifiers (red) are shown. Primers used for paired end (PE) or single end (SE) sequencing workflows are indicated.

| Name | Sequence (5' – 3') | Used for |
|---|---|---|
| **Uni-F** | [Phos]gatcATCTCGTATGCCGTCTTCTGCTTG23ddC | Preparing Uni-DS, short A-T adaptor |
| **Uni-R** | CAAGCAGAAGACGGCATACGAGATgatcT | |
| **Uni-ext** | CAAGCAGAAGACGGCATACGAGATgatcT | Extension of different ends |
| **NL5-ext** | AATGATACGGCGACCACCGA | |
| **Long barcoded oligos for preparing long A-T adaptors with different barcodes (Bc)** | | |
| **NL5Aa-F** | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNTAAGCgatcT | Preparing NL5Bc-DS, A-T long adaptors with barcodes |
| **NL5Aa-R** | [Phos]gatcGCTTANNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT23ddC | |
| **Oligos for making splint adaptors by annealing to corresponding reverse ones above** | | |
| **NL5Aa-SL-F** | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNTAAGCgatcNNNNN | Preparing NL5Bc-SL, splint adaptor |
| **Uni-SL-F** | NNNNNAgatcATCTCGTATGCCGTCTTCTGCTTG23ddC | Uni-SL adaptor |
| **PCR-fw** | AATGATACGGCGACCACCGA | PCR primers |
| **PCR-rv** | CAAGCAGAAGACGGCATACG | |
| **Oligos for ChIP-exo-TT** | | |
| **Uni-dT(24)VN** | CAAGCAGAAGACGGCATACGAGATgatcTTTTTTTTTTTTTTTTTTTTTTTVN | Extension, SE |
| **PE-dT24VN-Bc** | AGACGTGTGCTCTTCCGATCTNNNNNNTAAGCTCAAGTGATCTTTTTTTTTTTTTTTTTTTTTTVN | Extension, PE |
| **GI-ext** | GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG | Extend poly-C |
| **PCR-fw** | AATGATACGGCGACCACCGA | PCR primer |
| **PCR-rv** | CAAGCAGAAGACGGCATACG | PCR, SE |
| **PE-Bc1-PCR-rv** | CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT | PCR, PE, fixed barcode in bold |
| **PCR-GI-rv** | GGCCACGCGTCGACTAGTAC | PCR GI/poly-C |
| **Oligos for SMART** | | |
| **PE-dT24VN-Bc** | AGACGTGTGCTCTTCCGATCTNNNNNNTAAGCTCAAGTGATCTTTTTTTTTTTTTTTTTTTTTTVN | Extension, PE, barcoded |
| **PE-dT24VN** | AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTVN | Extension, PE |
| **Btn-TSP-Bc** | [Btn]ACACGACGCTCTTCCGATCTNNNNNNTAAGCGATCACArGrGrG | Template switching |

| | | |
|---|---|---|
| **PE-Bc1-PCR-rv** | CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT | PCR primers with fixed barcodes in bold |
| **PE-Bc2-PCR-fw** | AATGATACGGCGACCACCGAGATCT**ACATCG**ACACTCTTTCCCTACACGACGCTCTTCCGATCT | |

### 2.2.17 SMART-DRIPc method for mapping R-loops

DRIPc experiment was carefully performed under RNase-free conditions. Stock Buffers were treated with 0.1% DEPC whenever possible, otherwise prepared with DEPC treated water. Similarly, fresh working solutions were prepared using DEPC treated water.

Yeast cell lysis and sonication steps prior to chromatin immunoprecipitation were done quite similar to those of ChIP-exo with few changes. Yeast cells were grown overnight in 250 µl YEA full medium to mid-log growth phase. Cells were washed twice with ice-cold 1X PBS without formaldehyde crosslinking then pelleted at 2800 xg for 3 min at 4 °C before moving to a screw cap tube, discarding supernatant and snap freezing in liquid nitrogen.

After thawing on ice, the cell pellet was dissolved in 750 µl ChIP lysis buffer (50 mM Hepes-KOH pH 7.5, 1 mM EDTA pH 8, 140 mM NaCl, 1% tritonX-100, 0.1% sodium deoxycholate, 1X FY protease inhibitor, 1 mM PMSF) and distributed equally to two 2-ml screw cap tubes filled with 750 µl cold zirconium beads. Bead beating was done using Precellys homogenizer following program; three times, 20 sec and 5500 rpm in a cold room with 2 min incubation on ice in between bead beating cycles. After making holes in the lysate-containing tubes, tubes were inserted in 5-ml glass tubes and centrifuged at 200 xg for 1 min at 4 °C to bring cells down. Cell lysate from each glass tube was again distributed equally to two 1.5-ml diagenode TPX tubes for sonication (each tube contained around 300 µl lysate). Sonication was done for 10 cycles at high power, 30 min on / 30 min off (10 min total) using diagenode Bioruptor. Sonicated chromatin samples from different tubes were mixed together to have an equivalent of two samples of two cultures. Mixed tubes were centrifuged at 3000 xg for 10 min at 4 °C to precipitate cell debris before moving sonicated chromatin to a fresh Eppendorf tube.

10 µg S9.6 antibody was added to each tube (two samples combined) before incubating overnight at 4 °C on a rotating wheel for chromatin immunoprecipitation. No-Ab control sample was included. Later, 80 µl of a 1:1 protein A/protein G beads mix prewashed thrice with FA lysis buffer were added to chromatin-antibody mix and incubated on a rotating wheel

at 4°C for 2 hours. After centrifuging chromatin-containing tubes at 94 xg for 1 min at 4 °C, supernatant was discarded and beads were washed twice with ChIP lysis buffer, twice with wash buffer I (50 mM Hepes-KOH pH 7.5, 1 mM EDTA pH 8, 140 mM NaCl, 1% tritonX-100, 0.1% sodium deoxycholate), twice with wash buffer II (50 mM Hepes-KOH pH 7.5, 1 mM EDTA pH 8, 500 mM NaCl, 1% tritonX-100, 0.1% sodium deoxycholate) and twice with wash buffet III (10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8, 250 mM LiCl, 1% NP40, 0.5% sodium deoxycholate), each with 1 ml buffer volume on a rotating wheel for 5 min (8 times total, 5 min each). Before the last wash, beads were resuspended in wash buffer III and moved to a fresh tube. After washing the beads for 5 min using 1 ml 1X TE (10 mM Tris pH 7.5, 1 mM EDTA), beads were equilibrated by rinsing with 1 ml of 10 mM Tris pH 7.5 before starting on-bead enzymatic reactions.

For RNase III treatment, beads (bed volume accounted for 40 µl) were incubated in 200 µl final reaction volume of 1X RNase III buffer, 1X MnCl2 and 10 U RNase III at 37 °C for 30 min on a rotating wheel. After washing the beads twice with wash buffer I and once with 1X TE then equilibrating with 10 mM Tris pH 7.5 as indicated above, P1 treatment reaction was done in 200 µl final-volume reaction containing 1X P1 reaction buffer and 25 U P1 enzyme at 37 °C for 30 min on a rotating wheel. After washing the beads of treated and non-treated samples, beads were diluted in 1 ml 1X TE buffer and split into two halves before performing the RNase H treatment. After equilibrating with 10 mM Tris pH 7.5, beads (20 µl bed volume) were incubated in 60 µl final volume reaction of 1X RNase H buffer and 25 U RNase H at 37 °C for 30 min.

Elution was conducted using 200 µl of SDS elution buffer (1% SDS, 25 mM Tris-HCl pH 7.5, 10 mM EDTA and 200 mM NaCl) containing 80 units proteinase K (Ambion) by shaking at 1000 rpm on a thermomixer at 50 °C for 30 min. Tubes were centrifuged at l35 xg for 1 min and liquid was moved to a fresh tube for phenol chloroform extraction which was done as indicated above for ChIP-exo experiments.

### 2.2.18 SMART library preparation for S9.6-immunoprecipiated RNA

Master mixes were prepared for all reactions whenever possible. Before library preparation, genomic DNA was eliminated from RNA using TURBO DNA-free™ Kit (Ambion, AM1907) following rigorous treatment conditions. In brief, RNA was treated with 1 U TURBO DNase in 50 µl reaction of 1X buffer and 40 U RNasin (Promega) at 37 °C for

37 min before adding another 1 U of DNase and incubating for further 15 min. Reaction was inactivated by adding 0.2 volumes of inactivation beads and incubating at 26 °C for 5 min with intermittent shaking. Tubes were centrifuged and liquid was moved to fresh tubes for RNA purification using Zymoresearch 25 kit.

RNA ends were dephosphorylated by treating with 1 U of rSAP (NEB; M0371S) in 15 µl reaction of 1X FS buffer and 20 U RNasin (Promega) at 37 °C / 15 min followed by inactivation at 65 °C for 5 min. RNA was A-tailed in a 17.5 µl reaction of 1X FS buffer, 1 µl of 3 mM ATP, 20 U RNasin (Promega), 5 U PAP (NEB; M0276S) at 37 °C for 30 min followed by inactivation at 65 °C for 20 min.

For reverse transcription and template switching, tailed RNA was incubated with 1 µl of 5 µM barcoded dT24VN primer at 65 °C for 5 min then snap chilled on ice before adding 1 µl of 5X first strand buffer, 1 µl of 10 mM dNTPs, 1 µl of 100 mM DTT and 20 U RNasin (Promega). Reaction components were incubated at 42 °C for 2 min then brought to room temperature before adding 1 µl of 200 U/µl SuperScript™ II (ThermoFischer Scientific; 18064014) for first strand cDNA synthesis at 42 °C. After 15 min, reaction tubes were brought to room temperature before adding 1 µl of 10 µM barcoded TSP primer and incubating again at 42 °C for an hour before denaturing at 70 °C for 15 min. *Taq* PCR check and final Q5 PCR amplification were done similar to ChIP-exo libraries, but without prior purification (2.2.19).

### 2.2.19 ChIP-exo libraries amplification and sequencing

Before the PCR check and library preparation, ligated DNA was purified using equal volume of AMPure XP beads (40 µl) according to manufacturer manual and resuspended in 20 µl 0.1X TE buffer. 2 µl of purified DNA was used as a template for PCR amplification with Promega GoTaq® DNA Polymerase (M3001) in a 20 µl reaction containing 1X green *Taq* buffer, 200 µM dNTPs, 0.5 µM of each primer, 1.25 units Promega *Taq* DNA polymerase. Hot start PCR amplification was done using program; 95 °C / 5 min, (95°C / 15 sec - 55 °C / 30 sec - 72 °C / 1:15 min) for 27 cycles, 72 °C / 5 min, 4 °C / ∞. Amplification product was run on 1% pre-stained agarose gel till clear resolution of bands. Unless stated otherwise, 1 µg of 2-log ladder (NEB) was used for PCR band size determination.

For final library amplification, half the purified DNA (9 µl) of selected libraries was used for the final PCR amplification using Q5® High-Fidelity DNA Polymerase (NEB M0491) in

25 µl reaction containing 1X Q5 buffer, 200 µM dNTPs, 0.5 µM of each primer, 0.5 units Q5 DNA polymerase. Hot start PCR amplification was done using program; 98 °C / 45 sec, (98°C / 10 sec - 72 °C / 1:15 min), 72 °C / 2 min, 4 °C / ∞. PCR cycles were speculated from the signal strength of the previous *Taq* PCR check. Amplified DNA was purified using equal volume of AMPure XP beads (40 µl) and resuspended in 20 µl 0.1X TE buffer. 1 µl of purified DNA was run on bioanalyzer to check fragment size distribution and concentration before mixing the samples at an equimolar ratio and sending for sequencing. List of amplified libraries and sequencing platforms is listed in Table 2.7.

**Table 2.7. List of sequenced libraries and sequencing platforms.**

| Factor | Cell type | Experiment | Seq. platform | Facility | Conc. |
|---|---|---|---|---|---|
| **R-loops** | *S. pombe* | ChIP-exo 1 | NextSeq, HT 75 SE | BRF, JCSMR | 2 nM |
| **R-loops** | *S. pombe* | ChIP-exo-TT | NextSeq, HT 75 SE | BRF, JCSMR | 2 nM |
| **H3K9me2** | *S. pombe* | ChIP-exo-TT | NextSeq, HT 75 SE | BRF, JCSMR | 2 nM |
| **R-loops** | *S. pombe* | SMART-DRIPc | Hi-seq 2000 PE | Novogene, Singapore | 3 nM |
| **Kre33-FTP** | *S. cerevisiae* | ChIP-exo-TT | Hi-seq 2000 PE | Novogene, Singapore | 3 nM |
| **H3K9me2** | U2OS | ChIP-exo-TT | Hi-seq 2000 PE | Novogene, Singapore | 3 nM |
| **AhR** | Mouse EL4 | ChIP-exo-TT | Hi-seq 2000 PE | Novogene, Singapore | 3 nM |
| **AhR** | Mouse EL4 | ChIP-exo-SMART | Hi-seq 2000 PE | Novogene, Singapore | 3 nM |

## 2.3 *In vitro* transcription and preparation of artificial R-loops

For artificial R-loops preparation, two single stranded DNA fragments of 500-nt length and *in vitro* transcribed RNA were used. The ssDNA fragments are complementary to each other at 190 nt from each end, while the middle 120 nt are non-complementary. The RNA used was complementary to the middle 120 nt of the reverse DNA strand. This reverse strand which belongs to mammalian actin-B terminator was used as a template for amplification using Phusion polymerase to amplify the middle 120 nt sequence and abend the T7 promoter sequence for RNA synthesis by *in vitro* transcription.

500 ng of amplified B-actin fragment was used for *in vitro* transcription in 20 µl reaction containing 1X Hi-T7 RNA polymerase buffer (NEB), 1 mM rNTPs, 20 U RNasin (Promega) and 20 U Hi-T7 RNA Polymerase. Reaction mix was incubated at either 37 or 50 °C for 2 hours followed by DNase treatment and RNA extraction using Zymoresearch 25 kit. 1 µg of purified RNA was checked by separation on 2% agarose gel. Before loading on gel, RNA was

denatured in 20 µl denaturing buffer (1% Formaldehyde, 60% formamide) by incubation at 72 °C for 5 min then snap chilling on ice. 1 µg of Riboruler low-range RNA ladder (ThermoFischer Scientific) was loaded as a marker.

100 ng of each DNA and 1 µg of RNA were used for hybridization. The three strands (two ssDNA and *in vitro* transcribed RNA were annealed in 40 µl 1X annealing buffer (10 mM Hepes-KOH pH 7.5, 0.1 mM EDTA, 50 mM NaCl) by heating in a thermoblock at 72°C for 5 min and switching off for slow cooling. Annealing products were checked using agarose gel electrophoresis and dsDNA and ssDNA controls were loaded in parallel. Half the annealing reaction (20 µl) was used for nuclease P1 treatment reaction to verify the annealing product. P1 treatment was done in 26 µl volume containing 1X NEB1 buffer, 1X BSA and 0.5 U P1 (NEB).

**Table 2.8. Sequences of primers and DNA constructs used for generation of artificial R-loops.** Middle R-loop-forming non-complementary sequence (red) and T7 promoter sequence (yellow) are indicated.

| Fragment | Sequence |
|---|---|
| **R-loop random fw** | CTCGGGGCCAGATTTAGGGTGAGCTCTTCCGTGATAAAGACTGTAAACCTCCCCTTACACCCAGAG GGTGCTGATCAGGCTCAGGGTTGCTTATTCCAGTGAGGGAGGATGACGGACGAAGCGTTCTGTCG ACGCAGGAATAGCTAAAGTCTGTATACCCTTTGCACGTCGCTTGTCCTACTGGGCGTTACAATTATT CCGCATGATCGGGTAGAAACGCGATAGTCTAACTTAGGCAAAGTAAGATACAGAGGAGACATTGA CCAGGGGCAGCTGGCCTAGTCGGCGATACCCACTTGAAGATACCGAGCATTACGTCTTCCTTACCG AGGAGAACGCGTGACTTTCGTGACGCACCGCTCGCGGGAATAGCATGAGGGAGGTATAGATTTGA TGCTCCCAAAAAACCAGAGGCAACAAATTGGATGTGCTCTGGGTCAAAATGCACATGAGCCGTCCG TTTTCACAGCGTCCTATTAATACCAGACAGGGTTTGGAGG |
| **R-loop actin rv** | CCTCCAAACCCTGTCTGGTATTAATAGGACGCTGTGAAAACGGACGGCTCATGTGCATTTTGACCCA GAGCACATCCAATTTGTTGCCTCTGGTTTTTTGGGAGCATCAAATCTATACCTCCCTCATGCTATTCC CGCGAGCGGTGCGTCACGAAAGTCACGCGTTCTCCTCGGTAAGGAAGACGTAATGACCCCCACTCA AGGGAACACCGTGCCCAGCTACCCTGAGACAGCCCCACTCCAGGAAATGCAGGTGCCAACCAGCC CCAGTGAGGCATGGCACCTGAGCCAGACACCCCCAAATAGTCCCTAACGCCCAGTAGGACAAGCG ACGTGCAAAGGGTATACAGACTTTAGCTATTCCTGCGTCGACAGAACGCTTCGTCCGTCATCCTCCC TCACTGGAATAAGCAACCCTGAGCCTGATCAGCACCCTCTGGGTGTAAGGGGAGGTTTACAGTCTT TATCACGGAAGAGCTCACCCTAAATCTGGCCCCGAG |
| **PCR.actB.rv** | ACCCCCACTCAAGGGAAC |
| **IVT.actB.fw** | TAATACGACTCACTATAGGGAGAGGGACTATTTGGGGGTGTC |

## 2.4 Bioinformatic analysis

Illumina reads were subjected to thorough quality control steps; weak quality bases were trimmed from the 3' end of the reads with a sliding window strategy using Trimmomatic (with the parameter SLIDINGWINDOW : 4 : 20) **(Bolger *et al.*, 2014)**; the maximum allowed Dust score (a measure of low complexity ranging from 0 to 100) for the sequences was 7; and, finally, reads shorter than 38 nt were filtered out from the libraries. QC filtered reads from Total mRNA-seq, NET-Seq and R-loop libraries were aligned using hisat2 **(Kim *et al.*, 2019)** to the pombe genome (ASM294v2, PomBase), not allowing splicing for R-loop libraries (with parameter --no-spliced-alignment). For mRNA-seq libraries, 2000 nt was set as the maximum allowed intron length.

Overlaps among regions were analysed with BEDTools **(Quinlan & Hall, 2010)** and Venn diagrams were visualised by VennMaster **(Kestler *et al.*, 2005)**. R-loop positive regions were predicted with MACS2 using the narrow peaks option **(Y. Zhang *et al.*, 2008)**, separately, from the forward and reverse strand. Bigwig tracks were generated using bamCoverage from deepTools **(Ramírez *et al.*, 2016)**; –filterRNAstrand option was set based on the orientation of Read1 from the corresponding library. In the literature RNA-seq, NET-seq and DRIPc libraries, Read1 aligns on the 3'end of the fragment (opposite strand) while the second mate is on the 5' end (same orientation). In contrast, in SMART-DRIPc libraries, the orientation of Read1 is the same as the gene orientation representing the RNA component of the RNA-DNA hybrid. Check https://deeptools.readthedocs.io/en/develop/content/tools/bamCoverage.html for futher details.

Bigwig tracks were then visualised in IGV **(Robinson *et al.*, 2011; Thorvaldsdóttir *et al.*, 2013)**. The union of these predictions was then ranked by the forward signal and visualised using computeMatrix and plotHeatmap from deepTools2. Metagene plots for R-loop maps were done by ngsplot **(Shen *et al.*, 2014)**. Spearman-based correlation analysis of rRNA and tRNA loci was performed in R using the ggplot2 package **(Wickham, 2016)**.

# Chapter 3. Development of High-sensitivity ChIP-exonuclease (hsChIP-exo) Methods

## 3.1 Introduction

Mapping protein-DNA interaction for defining genomic location and enrichment of different proteins is a critical mean for understanding the functions and modes of regulation of these proteins and, also, for uncovering possible interactions with other factors and structures. Chromatin immunoprecipitation (ChIP) is a powerful technique for mapping different DNA-binding proteins, transcription factors and histone marks. Importantly, it provides insights about the DNA binding sequences of these factors. However, contamination from unbound DNA and lack of resolution can sometimes represent drawbacks for defining the exact binding sequence. The elegant ChIP-exonuclease (ChIP-exo) technique, a modification of the conventional ChIP, was introduced to overcome these obstacles and boost the ChIP power. The hallmark of the new technique is the incorporation of a lambda exonuclease treatment step to digest 5' ends of dsDNA to the site of protein-DNA crosslinking. After DNA sequencing and adaptor trimming, 5' ends of sequenced reads are mapped to the genome to capture the forward and reverse 5' borders of protein-DNA crosslinking site. This was found to reveal the exact DNA binding sequence for different proteins and render a near single-nucleotide resolution mapping **(Rhee & Pugh, 2011**, **2012)**.

As shown in Figure 3.1A, ChIP-exo method starts similar to the classic ChIP technique. Cells are fixed for protein-DNA interactions to be crosslinked before cell lysis and homogenization. Chromatin is then sonicated to generate a desirable DNA fragment size before immunoprecipitation. Immunoprecipitation is performed by incubating with the required antibody which binds target immunocomplexes to be later pulled down with affinity beads. Finally, affinity beads are moved to fresh tubes and stringently washed to remove the background. For ChIP-exo, multiple enzymatic treatments are carried out while the chromatin/DNA is still bound to the beads (on-bead reactions). DNA ends are first blunted then 5' ends are phosphorylated. 3' ends of DNA are dA-tailed prior to ligating double stranded adaptor with dT overhang (3' end adaptor). After the first ligation (3' end ligation), the DNA is phosphorylated again to be treated with lambda exonuclease that digests 5' ends of DNA to

the site of crosslinking. Another enzymatic treatment with Rec$_{Jf}$ exonuclease is also included to digest single stranded DNA that may non-specifically bind to the beads, before decrosslinking and elution are done. After DNA extraction and precipitation, enzymatic reactions (in-solution reactions) are carried out toward a second ligation step (5' end ligation). DNA is denatured for an extension primer to be annealed to 3' ends of DNA. This primer is used for an extension step for generation of second DNA strand and formation of dsDNA. After a second dA-tailing step, double stranded adaptor with dT overhang (5' end adaptor) is ligated to dsDNA. Finally, the manipulated DNA is amplified and sequenced.

Despite its advantages, ChIP-exo workflow is quite long and laborious, requiring more than five days of continuous work and careful conduct. In addition, it includes multiple on-bead and in-solution enzymatic treatments interspaced with several washing and purification steps, respectively. These incubations and extensive washing steps may lead to depletion of low-abundance target interaction sites, thereby affecting the final signal enrichment. For these reasons, the use of ChIP-exo has been quite limited. Another version of ChIP-exo called ChIP-nexus was introduced to reduce enzymatic treatment steps and improve ligation efficiency. The hallmark of the ChIP-nexus is the use of the ssDNA circularization principle of the iCLIP methodology **(J. Konig *et al.*, 2010)**. This step comes after DNA blunting, first adaptor ligation and lambda exonuclease treatment **(He *et al.*, 2015)**.

Similar to the original ChIP-exo (also called ChIP-exo 1), ChIP-nexus technique is also time consuming and complicated to conduct. Recently, different simplified versions of ChIP-exo have been introduced to enhance the sensitivity and decrease the hands-on time. The major modification was to abandon the double stranded A-T ligation and adopt splint ligation as the later requires less enzymatic treatments. The new versions were called ChIP-exo 4 and ChIP-exo 5 (Figure 3.1B, C). ChIP-exo 4 uses two splint ligations with the exonuclease treatment before both of them. Differently, ChIP-exo 5 maintained the first A-T ligation followed by exonuclease treatment steps in an order similar to the original ChIP-exo 1 protocol, but ends with a second splint ligation. According to the authors, these changes, greatly, enhanced the resolution and shortened the time of the experiments **(Rossi *et al.*, 2018)**.

Here, I genuinely sought to deploy the original ChIP-exo method to capture H3K9me2 histone mark in fission yeast. I found that this mark is hard to detect in this organism and requires a higher-sensitivity approach. Therefore, I tried the two simplified versions, ChIP-exo

4 and ChIP-exo 5, to find that they are less robust than the original one. Given these findings, I mainly aimed to develop higher-sensitivity ChIP-exo methods (hsChIP-exo) that also maintain the high-resolution of original ChIP-exo for mapping H3K9me2. I hopothesized that replacing A-T ligation with a ligation-free method would enhance the sensitivity of ChIP-exo and enable the capture of H3K9me2 histone mark in fission yeast. First, in an effort to reduce on-bead enzymatic reactions and enhance sensitivity, I replaced the first A-T ligation with poly dA-tailing in a method that I called ChIP-exo-TT. I found ChIP-exo-TT method to be more sensitive than the original ChIP-exo 1. With ChIP-exo-TT, I mapped H3K9me2 in fission yeast and in low-input mammalian samples. I also used it to capture Kre33, a single-locus DNA-binding protein, in budding yeast and ligand-activated AhR, a low-abundance transcription factor, in mammalian lymphocytes. This couldn't have been possible using the ChIP-exo 1.

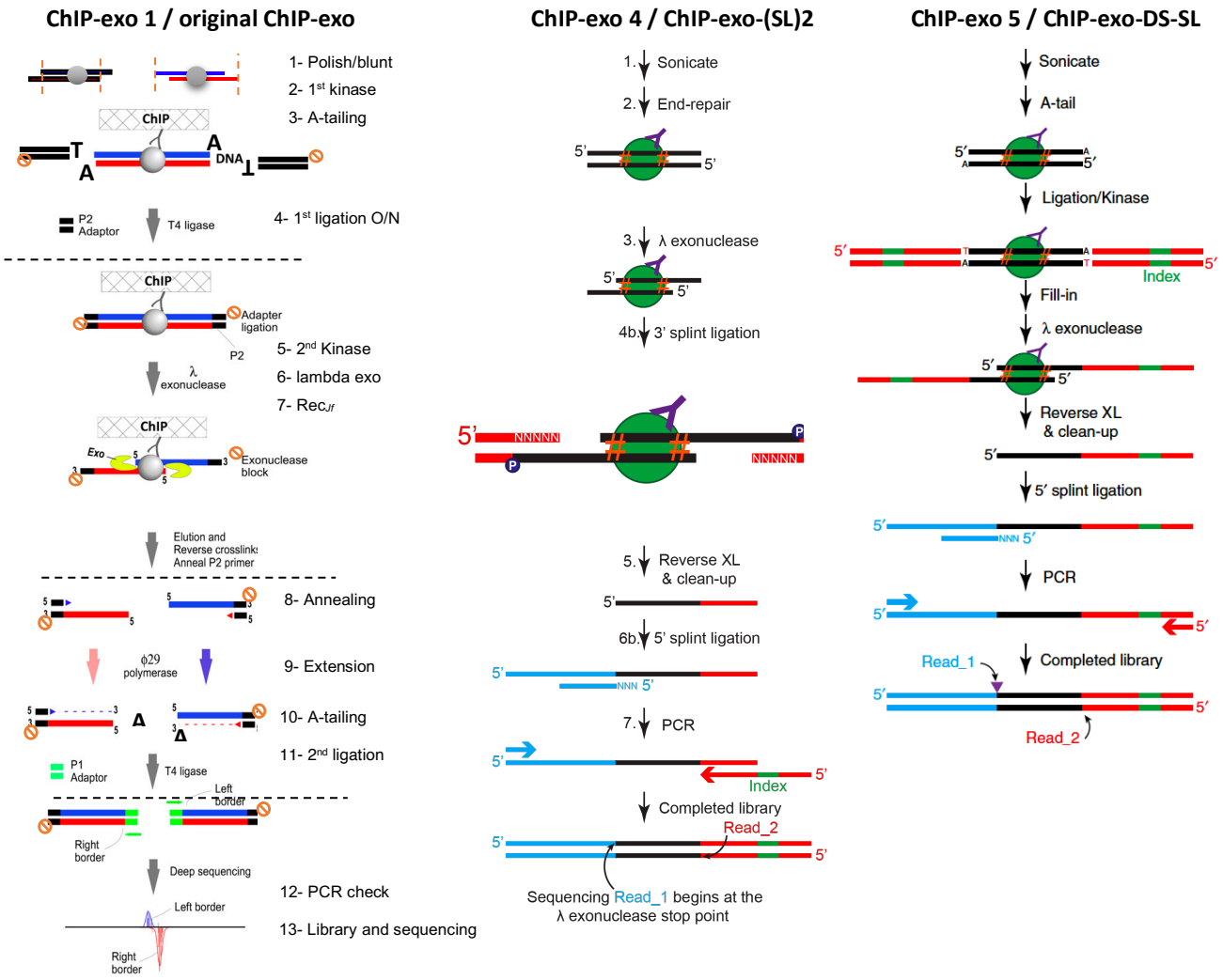



**Figure 3.1. Outlines of ChIP-exo 1, 4 and 5 methods workflows.** See text for details. Adapted from **Rhee and pugh (2011)** and **Rossi *et al*. (2018)**.

In a continuing effort to reduce the number of on-bead and in-solution reactions and to improve the sensitivity of ChIP-exo, I adopted the SMART technology for ChIP-exo DNA that was poly-dA tailed as in ChIP-exo-TT, in a separate method that I called ChIP-exo-SMART. SMART technology is a robust ligation-free method that is based on cDNA synthesis and template switching using a dedicated reverse transcriptase. These improvements reduced the hands-on time to 2 hours roughly and shortened the total working time by two days. ChIP-exo-SMART showed around ten- and two-fold higher PCR amplification sensitivity compared to the original ChIP-exo and ChIP-exo-TT protocols, respectively. Unfortunately, it tended to enhance background noise and mask positive signals due to an overamplification drawback.

**Table3.1. All steps toward different versions of ChIP-exo method.** Different ChIP-exo versions used in this study are listed according to the degree of similarity. Different enzymes and primers used for comparable steps are indicated. The upper part of the table above the yellow row represents the on-bead reactions while the lower part represents the in-solution reactions. Similar colours represent similar steps. Reactions used without difference from the original ChIP-exo are shaded in yellowish green and green.

| Version<br>Reaction | ChIP-exo 1 | ChIP-exo-DS-SL | ChIP-exo-(SL)2 | ChIP-exo-SL-DS | ChIP-exo-TT | ChIP-exo-SMART | Low input ChIP-exo-SMART |
|---|---|---|---|---|---|---|---|
| **Blunting**<br>**1st kinase** | Separate or combined | Separate or combined | Separate or combined | Separate or combined | Combined or omitted | | |
| **dA-tailing** | dA-tailing | dA-tailing | | | | | |
| **1st ligation** | A-T, Uni-DS | A-T, Uni-DS | | | | | |
| **2nd Kinase** | Kinase | Kinase | | | | | |
| **Lambda exo** | Lambda exo | Lambda exo | Lambda exo | Lambda exo | Lambda or T7 exo | T7 exo | T7 exo |
| **Rec*Jf* exo** | Rec*Jf* exo | Rec*Jf* exo | Rec*Jf* exo | Rec*Jf* exo | | | |
| | | | Splint ligation, NL5-SL | Splint ligation, NL5-SL | | | |
| **Elution, RNA digestion, decrosslinking, and DNA extraction and precipitation** | | | | | | | |
| | | | | | | | SAP |
| | | Splint ligation, Uni-SL | Splint ligation, Uni-SL | | Poly-dA/dC tail | Poly-dA tail | Poly-dA tail |
| | | | | | 1.8 x AMPure beads | | 1.8 x AMPure beads |
| | | | | | | SMART | SMART |
| **Denaturing, annealing** | Denaturing, annealing | | | Denaturing, annealing | Denaturing, annealing | | |
| **Extension primer & Enzyme** | Uni-ext<br>Phi29 | | | NL5-ext<br>Phi29 | dT24VN or GI<br>Phi29 or Bst | | |
| **dA-tailing** | dA-tailing | | | dA-tailing | dA-tailing | | |
| **1.4 x AMPure XP beads** | | | | 1.4 x AMPure XP beads | | | |
| **2nd ligation** | A-T NL5-DS | | | A-T NL5-DS | A-T NL5-DS | | |
| **1 x AMPure XP beads purification** | | | | | | | |
| ***Taq* Check & Q5 PCRs** | PCR-fw & PCR-rv primers | | | | PCR-fw & PE-Bc1-PCR-rv | PE-Bc2-PCR-fw & PE-Bc1-PCR-rv | |

## 3.2 Results

### 3.2.1 ChIP-exo 1 works robustly for abundant proteins, but can't capture H3K9me2

Steps toward the different versions of ChIP-exo method are outlined in Table 3.1 highlighting varying conditions, enzymes and primers used for each version. I have found that all the steps count toward the success of the ChIP-exo experiment, starting from the crosslinking to the final library preparation. Efficient lysis, especially for yeast cells is one of the steps that may be underestimated, but is critical for the success of the experiment. A fast lysis under controlled temperatue is critical to avoid heating the sample and denaturing sensitive enzymes. Among multiple devices, I found Precellys device to be a good fit for efficient yeast cell lysis following program; 3 x 20 sec at 5500 rpm with 1-2 min break on ice.

Sonication reproducibility and DNA fragment size are two related factors that I found crucial for the reproducibility of the results. All sonications toward ChIP-exo experiments were done using Covaries S2 device which produced an average fragment size of 200 bp that was very consistent and reproducible all the times using parameters; duty factor (DF) 20; Intensity (I) 5; and cycles per burst (CPB) 200 for 8 min (Figure 3.2A, B and Figure S1). I tried to use Covaries ME220 for its its new automated options, but the sonication wasn't reproducible for fission yeast (Figure S1, all conditions and results are included).

One of the great advantages of the ChIP-exo over the classic ChIP is the decreased background noise to signal ratio. As a rule, to check the background contamination level, a negative control is always included to define the false-positive loci and exclude them from the result. The source of this background noise is the nonspecific DNA binding to either the beads used to pull down the antibody-immunocomplexes or the antibody itself. Except for some special cases, such as mapping FTP-tagged proteins, a no-antibody sample is used as a negative control to confirm specificity of the mapping. In my hands, I found that using either protein A beads alone or a mix of protein A and protein G beads clearly eliminated the background signal that the negative control samples were almost empty and couldn't be used for library amplification and sequencing. However, I found that using protein G beads for pulling down some rat antibodies, of IgG isotypes that has affinity only to protein G, produces high background. This can be clearly shown by comparing the PCR signals of the no-antibody libraries of Figure 3.2A, D and Figure 3.2C.

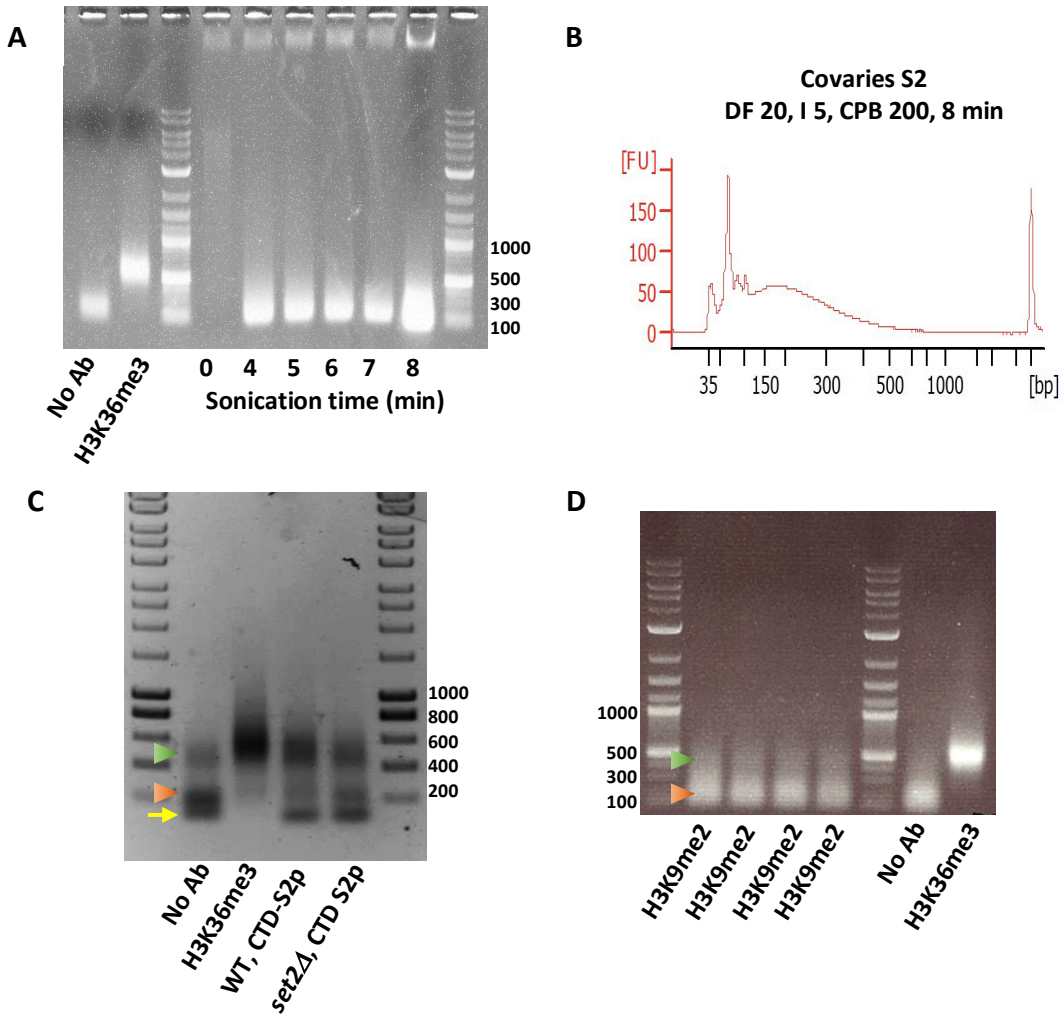



**Figure 3.2. ChIP-exo 1 is robust for abundant proteins and histone marks but not for H3K9me2. (A)** agarose gel analysis for PCR-amplified ChIP-exo 1 libraries (left two lanes) showing robust amplification for positive H3K36me3 library, but empty no-antibody control. DNA sonicated for different time-points was run in parallel to check fragment size and sonication efficiency. **(B)** bioanalyzer analysis for DNA sonicated using Covaries S2 device using the indicated parameters, showing DNA of ~150 - 200 bp average fragment size. **(C)** agarose gel separation for different ChIP-exo 1 libraries showing successful amplification for libraries compared to the no-Ab control which has weak signal (green arrow head), strong adaptor dimers (orange arrow head) and primer dimers (yellow arrow). **(D)** agarose gel analysis showing no amplification for H3K9me2 libraries using ChIP-exo 1. Protein A/G mix was used for all experiments except in **(C)** where only protein G beads were used (compare signals of the no-Ab controls). 1 µg of 1Kb Hyperladder (Bioline) was used in **(C)** while 2-log ladder (NEB) was used in **(A)** & **(D)**.

Unfortunately, there is no way to check the success of each single reaction before the *Taq* PCR check, the only quality control which is carried out at the end to check the success of the experiment in general. I usually carry out this PCR for 27 cycles in order to check the enrichment and quality of the libraries after loading on agarose gel. A positive good-quality library shows a band around 500 bp mark with minimal adaptor dimers while an empty library produces only adaptor dimers extending between 100 to 200 bp (Figure 3.2). This check is also

required to define the number of PCR cycles required for the final Q5 amplification toward sequencing. Importantly, minimal cycling is recommended to avoid generating PCR duplicaes. Based on the amount of input DNA, a cycle number ranging from 12 to 25 can be used **(Rhee & Pugh, 2011**, **2012)**. In this study, I carried out the final amplification using cycle number starting from 14 and not above 20 (for poor libraries). After the final amplification, the library is run on bioanalyzer to check the DNA size distributaion and confirm the absence of overamplification signs and PCR trailing due to ssDNA.

PCR amplification for different libraries processed by ChIP-exo 1 produced robust signals (Figure 3.2). Examples for antibodies used for immunoprecipitation and ChIP-exo 1 experiments are antibodies specific for histone H3, H3K36me3 histone mark, or for different phosphorylated CTD residues of RNA Pol II Rbp1 subunit (Figure 3.2A, C & D). A clear example for a successful ChIP-exo 1 and high-quality libraries will be shown in the next chapter for R-loop mapping. Disappointingly, after several trials to capture H3K9me2 in fission yeast, I couldn't have enough DNA for library preparation and sequencing (Figure 3.2D).

### 3.2.2 H3K9me2 mark is hard to detect in fission yeast using Western blotting

In parallel to ChIP-exo, I tried to directly detect changes in the abundance levels of H3K9me2 histone mark in *S. pombe* in response to specific molecular cues using Western blotting. After several Western blot experiments, I couldn't detect any signal in fission yeast, so I decided to use different batches of H3K9me2 antibody along with H3K36me3 antibody as a positive control for fission yeast. I also used a mammalian cell line as a positive control for H3K9me2 mark. Indeed, I successfully detected strong signals for H3K36me3 in yeast and H3K9me2 in mammalian cells, but not in fission yeast (Figure 3.3). Consistent with previous reports **(Cam & Whitehall, 2016; K. Zhang *et al.*, 2008)**, my observations confirm that H3K9me2 histone mark is hard to detect in fission yeast and may require a high sensitivity method for genome-wide mapping.

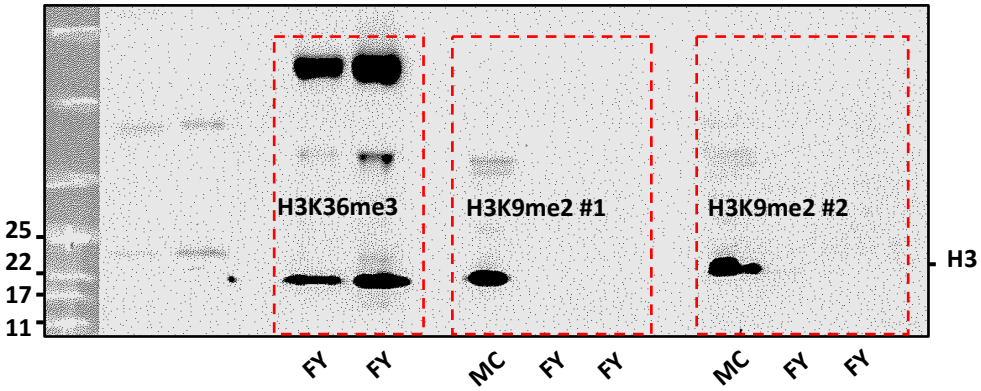



**Figure 3.3. H3K9me2 histone mark is hard to detect in fission yeast by Western blotting**. Semi-dry Western blot analysis showing robust signals for H3K36me3 in fission yeast (FY) and strong signals for H3K9me2 in a mammalian cell line (MC), but not in fission yeast. Two different batches of H3K9me2 antibody (#1 and #2) were used in parallel to detect H3K9me2 in fission yeast and in a mammalian cell line. H3K36me3 Ab was used as a positive control for a histone mark in fission yeast. Anti-rat and anti-rabbit secondary antibodies were used after H3K9me2 and H3K36me3 respectively. Primary antibody was used at 1:1000 dilution while secondary antibody was used at a 1:2000 dilution.

### 3.2.3 ChIP-exo-(SL)2 with two splint ligations is less efficient than ChIP-exo 1

The major change in the recently introduced ChIP-exo versions, compared to original ChIP-exo 1, is the replacement of double stranded A-T ligation by splint ligation. Splint ligation refers to the ligation of a splint adaptor (double-stranded DNA adaptor with an overhang of random nucleotides at one of the ends) to DNA with single stranded ends. I sought to develop and apply these new versions for mapping low-abundance DNA-binding proteins and histone marks such as H3K9me2 in fission yeast. Reportedly, ChIP-exo 4 and ChIP-exo 5, greatly reduced the hands-on time to half that of ChIP-exo 1 and maintained the high resolution of the ChIP-exo, so I decided to apply these two versions. Hereafter, I will call ChIP-exo 4 and ChIP-exo 5 as ChIP-exo-(SL)2 and ChIP-exo-DS-SL, respectively, due to the fact that ChIP-exo 4 has two splint ligations (SL)2 while ChIP-exo 5 has a double stranded (DS) ligation followed by a splint (SL) ligation step.

#### *3.2.3.1 First trial*

I chose to start with the ChIP-exo-(SL)2 version as the simplest one. Similar to original ChIP-exo 1, multiple reactions were applied for the bead-bound chromatin-DNA complexes. First of all, I applied no changes for DNA blunting and phosphorylation which I have done quite similar to ChIP-exo 1, while the next dA-tailing step was omitted as there is no need for

it for the splint ligation. For this method, exonuclease digestion step is done directly after the blunting and phosphorylation and before both ligation steps (Figure 3.1). For the exonuclease treatments, I used tritonX-100 and DMSO following the published protocol **(Rossi *et al.*, 2018)**. However, I found that this leads to the dispersion of the beads on the walls of the tubes which might lead to loss of the samples. I did the splint ligation directly after the exonuclease treatments to ligate splint adaptors to the single stranded 3' end of DNA. After elution and DNA extraction, I did the second ligation to ligate different splint adaptors to the 5' end of DNA. The check PCR generated only strong adaptor dimers without any amplification bands.

#### *3.2.3.2 Second trial*

For the next trial, I abandoned the use of DMSO and tritonX-100 for the exonuclease steps and turned back to use the regular reaction volumes and components of the original ChIP-exo 1 protocol. I also set back to do ligations for long time overnight. Before the second ligation, I denatured the DNA to separate the strands and enhance the ligation by heating at 95 °C for 3 min then snap-chilling on ice before adding the ligation mix. Among the different trials, I tried to combine the two ligation reactions in one step after DNA extraction, but the experiment didn't work at all. The check PCR generated only strong adaptor dimers without any amplification bands.

#### *3.2.3.3 Third trial*

Due to the fact that I have applied changes only to ligation steps, but not the first blunting, first kinase, lambda exonuclease or Rec*Jf* reactions which I did similar to ChIP-exo 1, I started to think that there is a problem with the ligation steps. This may be due to the ligation conditions, components used or due to inefficiency of a preceding determining step. One important step that may affect the ligation is the 5' end phosphorylation before the second 5' end ligation. It's known that lambda exonuclease (done before phosphorylation) leaves phosphorylated 5' ends but this may have got compromised for a reason. In the next trial, I used conditions identical to the original ChIP-exo for the blunting, kinase, lambda exo and Rec*Jf* then I did the first splint ligation as in the second trial. Later, I set to focus on testing requirements for the second ligation using three conditions listed in Table 3.2. Since the DNA is supposed to be phosphorylated after lambda exonuclease treatment, I carried out the ligation without any prior phosphorylation in the first sample. It has been reported that both the phosphorylation and ligation reactions can be done in one step without prior heat inactivation

or purification **(Rossi *et al.*, 2018)**. Although this matches the trend for simplifying the protocol, reducing the time, and increasing the sensitivity, I had concerns toward this approach as the sticking of the kinase to DNA ends is known to hinder the ligase and decrease its efficiency. For this reason, I wanted to test this approach against the conventional strategy of sequential phophorylation ligation reactions. Accordingly, for the second sample, I did the phosphorylation first then heated the sample to 95 °C for quick inactivation before starting ligation in the same mix, while for the third sample, I started both reactions in one mix instantly.

**Table 3.2. Different conditions used for 2nd kinase and 2nd splint ligation reactions of ChIP-exo-(SL)2.** For sample 1, PNK wasn't added in the reaction, DNA was heated at 95 °C for 3 min then chilled on ice before adding the ligation reaction mix. For sample 2, DNA was heated at 95 °C for 3 min then chilled on ice before adding the whole kinase ligase reaction mix, and the sample was left at room temperature for 30 min to enhance the kinase reaction then kept at 16 °C overnight for ligation. For samples 3 & 4, DNA was heated at 95 °C for 3 min then chilled on ice before adding the kinase reaction mix, but without adaptors or ligase. The reaction was incubated at 37 °C for 30 min. Afterwards, the DNA was heated again at 95 °C for 3 min then chilled on ice before adding the adaptors and the ligase and finally incubated at 16 °C overnight in a PCR machine. All volumes are in µl.

| Samples / Components | 1 | 2 | 3 | 4/No-Ab | Final conc. |
|---|---|---|---|---|---|
| **DNA** | 20 | 20 | 20 | 20 | |
| **H2O** | 11.5 | 9 | 9 | 9 | |
| **10X T4 DNA ligase buffer** | 4 | 4 | 4 | 4 | 1X |
| **10 µM Uni-SL adaptor** | 1.5 | 1.5 | 1.5 | 1.5 | 0.375 µM |
| **400 U/µl T4 DNA ligase** | 3 | 3 | 3 | 3 | 1200 U |
| **10 U/µl T4 PNK** | - | 2.5 | 2.5 | 2.5 | 25 U |
| **Final volume** | 40 | 40 | 40 | 40 | |

As expected, the first sample (no Kinase) showed the best amplification cofirming that DNA ends are already phosphorylated after lambda exonuclease treatment and compatible for ligation. The second sample showed the weakest amplification confirming my concerns about the instant kinase/ligase strategy. Unfortunately, check PCR produced weak amplification signals and strong adaptor dimers for all the samples (Figure 3.4A). This result wasn't improved after multiple repeats using the best working conditions of the third trial (Figure 3.4B). These results suggest that splint ligation, carried out under currently tried conditions, is less efficient than A-T ligation, knowing that ChIP-exo-(SL)2 decreased the number of reactions and purifications times, and shortened the hands-on time compared to the original version.

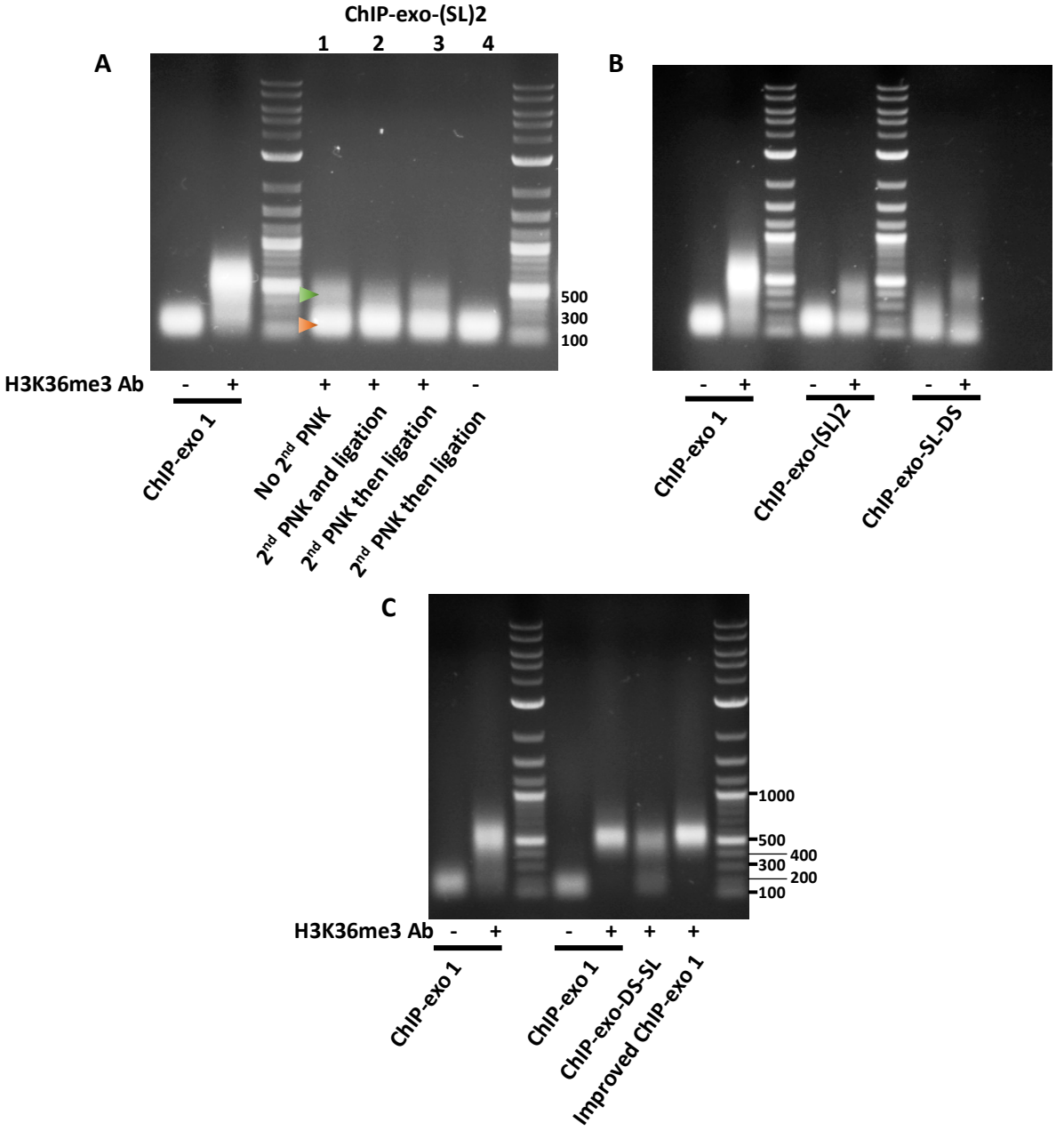



**Figure 3.4. ChIP-exo-(SL)2, ChIP-exo-SL-DS and ChIP-exo-DS-SL are less robust than ChIP-exo 1.** Agarose gel analyses for PCR-amplified libraries processed by multiple simplified versions of ChIP-exo, showing weaker signals (green arrow head) and stronger dimer contamination (orange arrow head) compared to ChIP-exo 1 libraries. **(A)** ChIP-exo-(SL)2; numbers represent the same samples listed in Table 3.3. **(B)** ChIP-exo-(SL)2 and ChIP-exo-SL-DS. **(C)** ChIP-exo-DS-SL, ChIP-exo 1 and improved ChIP-exo 1**.** H3K36me3 antibody was used for the optimization. PCR was done for 27 cycles, 20 µl were run on 1% agarose gel. 1 µg log2 NEB ladder was loaded as a marker. Refer to materials and methods for detailed protocols, PCR conditions and program.

### 3.2.4 The success of ChIP-exo experiment relies on the efficiency of the first ligation

Beside ChIP-exo 1 and ChIP-exo-(SL)2, I tried another version which is a combination of ChIP-exo-(SL)2 and ChIP-exo 1. More precisely, I did the first half of reactions (on-bead reactions) similar to ChIP-exo-(SL)2 toward a splint ligation. After DNA extraction, I did the in-solution reactions similar to ChIP-exo 1 with a primer extension step to a final A-T double stranded ligation. I called this version ChIP-exo-SL-DS. Similar to ChIP-exo-(SL)2, ChIP-exo-SL-DS showed weak signal and strong adaptor dimers (Figure 3.4B).

Next, I decided to swap the order of ligation steps, i.e., to do the double stranded before the splint ligation. For this, I did the on-bead reactions and double stranded A-T ligation down to the elution similar to ChIP-exo 1 protocol. After elution, I performed the reactions similar to ChIP-exo-(SL)2 ending by splint adaptor ligation. I called this method ChIP-exo-DS-SL. Surprisingly, the PCR check showed better amplification and weaker dimers contamination compared to the ChIP-exo-(SL)2 (Figure 3.4B, C). These findings confirmed my speculations that the splint ligation has less efficiency compared to the double stranded ligation using the corresponding conditions.

These results made me convinced that the original protocol, with two double stranded A-T ligations, is more efficient than the recent versions. Inspired by illumina protocols for DNA sample preparation, I tried to combine the first blunting and kinase reactions in one step. Excitingly, this increased the robustness of the final PCR signal (Figure 3.4C). For this reason, I decided to adopt these improvements for all the mapping experiments. These results confirm that the first ligation and the preceding reactions are determining steps for the success of ChIP-exo experiments. Unfortunately, these improvements still couldn't capture H3K9me2 in fission yeast.

### 3.2.5 Ligation-free Poly-dA tailing enhanced the sensitivity of ChIP-exo (ChIP-exo-TT) to capture H3K9me2 mark in fission yeast

To eliminate the need for long incubation times for adaptor ligation which is also dependent on preceding steps, I sought to replace adaptor ligation steps, especially the first one, with a ligation-free method. This tendency was inspired by others who used DNA tailing for robust and low-bias library preparation methods **(X. Peng *et al.*, 2015; Turchinovich *et al.*, 2014)**. I sought to adopt the same DNA-tailing principle for the ChIP-exo technique. First,

I carried out the lambda exonuclease step directly after blunting and kinase steps which I didn't apply any changes for. I eliminated the Rec$_{Jf}$ step as it was reported that inclusion or omission of this step does not make a difference **(Rossi *et al.*, 2018)**. After elution, I tried either poly-dA or poly-dC tailing for 3' ends of DNA using Terminal deoxy-Transferase (TdT) enzyme and either dATPs or dCTPs, respectively. After DNA purification using magnetic beads, I did primer extension with an oligo-dT or a guanosine-inosine (GI) primer to extend poly-dA or poly-dC tailed DNA, respectively, in order to generate dsDNA. Next, I performed the final dA tailing and A-T double stranded ligation. Hereafter, I will call this method as ChIP-exo-TT which stands for ChIP-exo using Terminal Transferase (Figure 3.5).

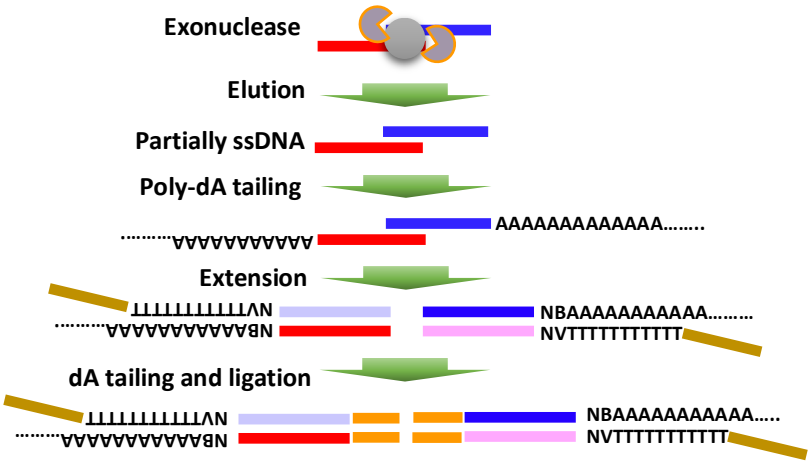



**Figure 3.5. ChIP-exo-TT method work flow.** Exonuclease treatment is done on beads (grey) to digest DNA 5' ends. After elution, DNA is poly-dA tailed for annealing of oligo-dT extension primer. After extension and dsDNA generation, DNA is dA tailed and double standed adaptor (orange) is ligated. For paired end sequencing, oligo-dT primer may contain a linker sequence (golden). poly-dA tailing is represented as the mainly used strategy.

Surprisingly, and for the first time, PCR amplification produced a strong signal for H3K9me2 immunoprecipitated DNA processed by this method (Figure 3.6A). Remarkably, the PCR signal was much stronger for the dA-tailed compared to the dC-tailed DNA. It's not very clear for me why the signal strength is different. This may be attributed to differences in the efficiency of the different tailing methods or in the ensuing annealing, extension and amplification steps. The reason for the later assumption is that different extension and amplification primers are used for the differently tailed DNA. Surprisingly, and in spite of the increased sensitivity of this method, there was no PCR signal in the no-Ab control sample compared to the positive sample processed by the same method (Figure 3.6).

To validate these results and confirm the specificity of the immunoprecipitation, I exploited the fact that the yeast cells have two copies for ura4 gene, a truncated endogenous copy (ura4-DSE) and a full copy ectopically inserted at the dg/dh heterochromatic pericentromeric outer repeat regions (otr-ura4). In principle, both copies should be amplified using the same primer pair to produce two variable-length amplicons specific for each allele. Surprisingly, the PCR amplification for the immunoprecipitated and exonuclease treated DNA (before further processing) produced one band specific for the full-length heterochromatic ura4 while the no-antibody library was devoid of any PCR amplification signals (Figure 3.6A). These findings strongly confirm the specificity and the validity of this method. Based on these promising results, I decided to adopt the poly-dA tailing strategy with which stronger amplification signal was reproducibly detected after ChIP-exo-TT (Figure 3.6B).

Next, I sought to apply some changes that would decrease the enzymatic steps, shorten the time and, more importantly, enhance the sensitivity of the whole method. I planned to replace lambda exonuclease with T7 exonuclease, a different exonuclease that doesn't require neither blunt nor phosphorylated ends. This means the blunting and kinase steps will be eliminated from the protocol. At the same time, I wanted to try different extension conditions before the A-T adaptor ligation. Beside regular extension with Phi29 followed by dA-tailing usign Klenow exo-, I planned to use Bst DNA polymerase, which has a dA-tailing activity, for a one-step extension & dA-tailing.

To achieve this, I did the immunoprecipitation for two samples using H3K9me2 and H3K36me3 antibodies, and included a no-Ab control sample. I processed the no-Ab and H3K9me2 samples similar to the above ChIP-exo-TT trial (polishing & kinase, lambda exo, elution, then poly-dA tailing to the end). For the H3K36me3 sample, I applied T7 exonuclease treatment as the only on-bead reaction before elution and DNA extraction. After the poly-dA tailing and DNA purification with magnetic beads, I split the DNA into two halves. The first half was extended with Phi29 DNA polymerase then dA-tailed with Klenow exo- as usual while the second half was extended and dA-tailed in one step using Bst DNA Polymerase (20 µl reaction volume containing 1X Thermopol buffer, 1 mM dNTPs, 1 µM extension primer and 8 U Bst DNA Polymerase large fragment at 60 °C for 1 hour) before the final ligation.

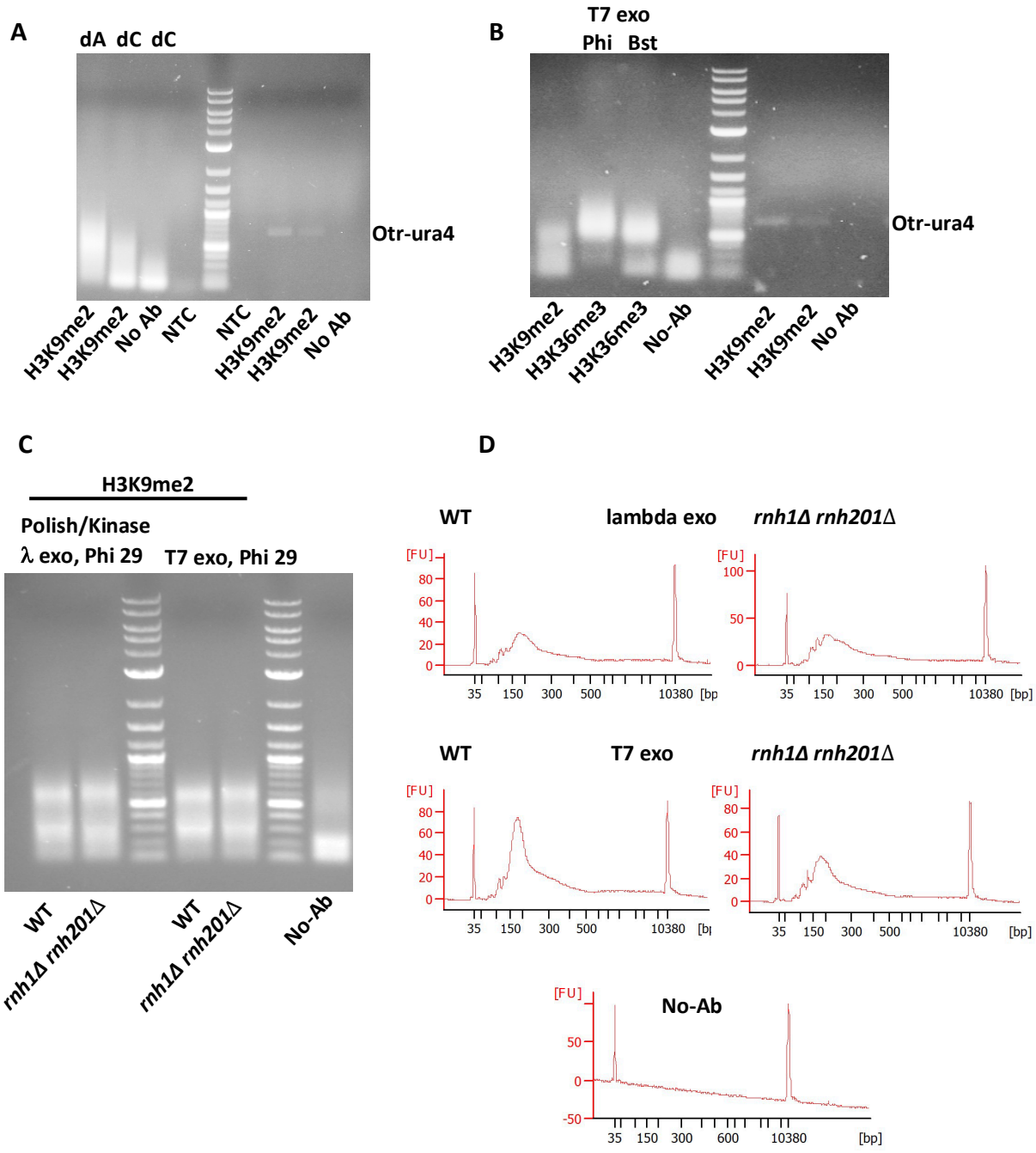



**Figure 3.6. ChIP-exo-TT is sensitive enough to capture H3K9me2 signal in fission yeast.** Agarose gel and bioanalyzer analyses for PCR-amplified libraries processed by ChIP-exo-TT with variable conditions. **(A)** check PCR signals detected for H3K9me2 libraries processed by ChIP-exo-TT using dA or dC tailing, while no-antibody and non-template control (NTC) are empty. PCR was done using 30 cycles. PCR validation produced only one signal specific for heterochromatin otr-ura4 (30 PCR cycles). **(B)** check PCR signals detected for H3K9me2 and H3K36me3 libraries processed with ChIP-exo-TT with dA tailing. One sample was used for H3K36me3 libraries, DNA was treated with T7 exonuclease then split into two halves before extension with either Phi29 or Bst polymerases. For ChIP-exo-TT, check PCR was done using 27 cycles while enrichment at ura locus was done using 32 cycles **(C)** check PCR signals for H3K9me2 mapping libraries for WT and *rnh1Δ rnh201Δ* fission yeast cells. ChIP-exo-TT was done following two work flows. PCR was done using 28 cycles **(D)** bioanalyzer profiles for libraries in **(C)** after final Q5 amplification using 20 cycles.

Reproducibly, the PCR amplification produced a decent signal for H3K9me2 library while the no-Ab was empty of any amplification signal except the adaptor dimers (Figure 3.6B). This strongly confirms that the ChIP-exo-TT is quite robust and can be used for mapping H3K9me2 histone mark and, most probably, other low abundance proteins. Excitingly, in spite of using half the amount of DNA after the immunoprecipitation for H3K36me3, very strong signals were still produced after PCR for each half. Comparing the signal for both of Phi29- and Bst-extended H3K36me3 DNA, the signal was much stronger and higher for Phi29-extended DNA. The reason for this difference may be attributed to either difference in the efficiency of annealing and extension, or final dA-tailing step. Regardless, I decided to adopt T7 exonuclease, with omitting the blunting & kinase step, for some experiments. I also continued using Phi29 and Klenow exo-, but excluded Bst DNA polymerase. With ChIP-exo-TT, I successfully mapped R-loops in fission yeast. I also mapped H3K9me2 mark in fission yeast and U2OS mammalian cells in the next chapters of the thesis.

### 3.2.6 ChIP-exo-TT captured single-locus DNA-binding protein kre33 in budding yeast

Excitingly, I got the chance to try the ChIP-exo-TT method for mapping different proteins in budding yeast *S. cerevisiae* and different mammalian cell lines. In budding yeast, for example, I studied the enrichment of FTP-tagged Kre33 as part of a collaborative project. Kre33 is an acetyltransferase required for acetylation of 18S ribosomal RNA in budding yeast **(Ito *et al.*, 2014; Sharma *et al.*, 2015)**. The main aim was to test whether this protein binds to ribosomal DNA repeats and whether there is a differential binding for the WT and mutant Kre33 protein to the 5′ External Transcribed Spacer (5’ ETS) of ribosomal DNA repeats. It was hypothesised that the mutant protein may have stronger binding to these regions. To test this hypothesis, I carried out a ChIP-exo-TT experiment for the different yeast strains based on FTP double affinity pulldown. On-bead reactions included only T7 exonuclease treatment. Poly-dA tailing was done after elution toward final ligation. Initially, the PCR check didn’t show strong amplification (Figure 3.7) which was also shown in bioanalyzer profiles. Surprisingly, however, after analyzing the data, the genome browser revealed a strong single signal, over chromosome 12, mapping to 35S ribosomal DNA locus in the WT and Kre33-mutant strain. Strikingly, the untagged control sample was clearly empty of any signal (Figure 3.7). These observations strongly evidence for the strong enrichment of this protein at these repeats. Amazingly, the genome browser signal exhibited an ultra-high resolution implying the efficiency of the T7 exonuclease.

Furthermore, genome browser view showed a generally stronger signal over 35S rDNA for the WT compared to the Kre33-mutant strain. Strikingly, the signal was roughly twofold stronger in the Kre33-mutant compared to the WT strain over the 5' ETS sequences (Figure 3.7). Consistently, qPCR analysis using primers that bind specifically to the 5' ETS of the 35S rDNA repeats, showed around l28 fold enrichment of WT Kre33 over 5' ETS sequence compared to the untagged control. Importantly, the same qPCR analysis revealed that mutant Kre33 has around twofold higher enrichment compared to the WT at the 5' ETS sequence. These observations strongly evidence that ChIP-exo-TT is a reliable method for mapping proteins with few DNA-binding sites. Mapping this kind of proteins is as hard as mapping low-abundance proteins as both bind to a small fraction of target DNA. Despite the abundance of Kre33 protein as being involved in ribosome biogenesis, its DNA binding sites are very limited which made it hard to map with previous ChIP-exo methods.

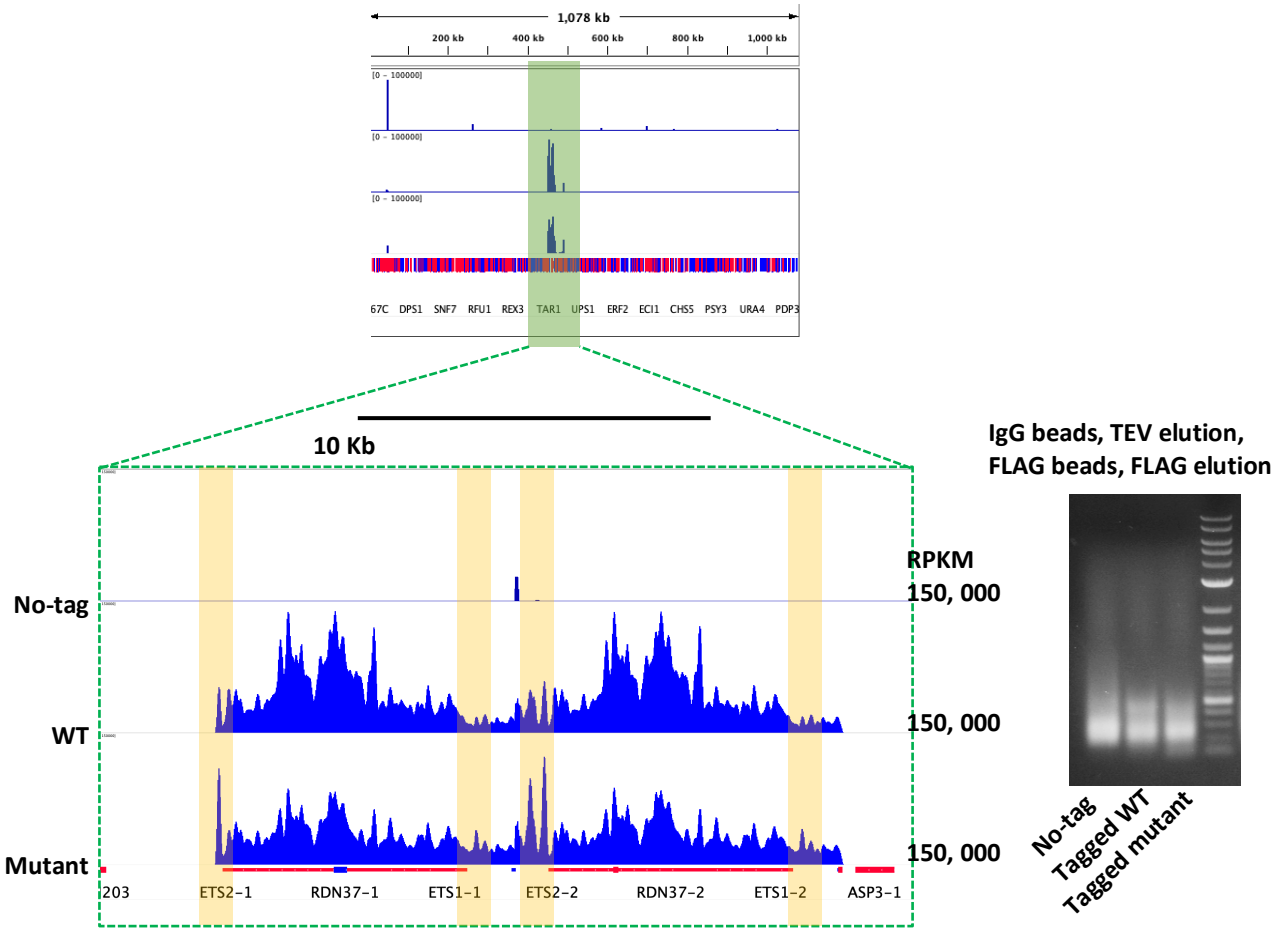



**Figure 3.7. ChIP-exo-TT detected a difference in enrichment of WT and mutant Kre33 protein over 5' ETS of rDNA in budding yeast.** Top; snap shot of genome browser for full chromosome 12, showing a single enriched region for 35S ribosomal DNA locus. Bottom left; zoomed-in genome browser view for enriched ribosomal DNA. While signal is generally stronger in WT compared to mutant cells, 5' ETS regions (yellow bars) show higher peaks in mutant compared to WT cells. Bottom right; agarose gel separation profiles after check PCR.

### 3.2.7 ChIP-exo-TT is compatible with low-input mammalian samples and captures ligand-activated transcription factors

Results obtained by ChIP-exo-TT for yeast Kre33 encouraged me to test this method for different mammalian cell lines and different proteins. To test this method for mammalian cells, I used either whole cell lysate or nuclear isolate of U2OS cell line as inputs for immunoprecipitation using H3K9me2 and H3K36me3 antibodies. Only 3 million cells were used for each immunoprecipitation sample. Amazingly, PCR amplification produced very strong signals for both different IP samples while the no-antibody control was clean (Figure 3.8B). These results confirm the versatility, high-sensitivity and specificity of ChIP-exo-TT. Next, I used this method to map H3K9me2 histone mark in WT and RNase H overexpression U2OS cell lines with only 300 k cells for the later.

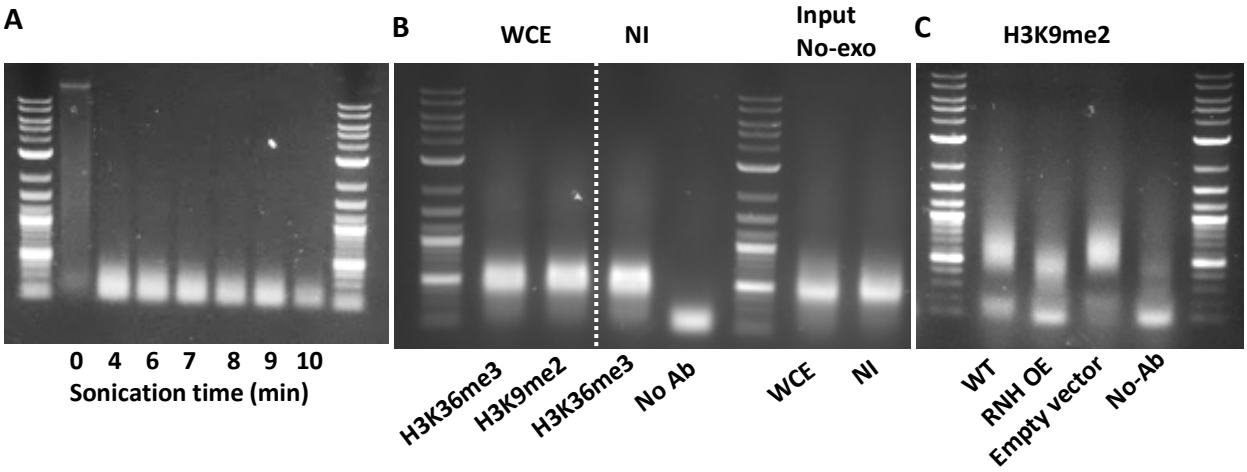



**Figure 3.8. ChIP-exo-TT is a versatile and high-sensitivity method compatible with low-input samples. (A)** agarose gel analyses for DNA sonicated at different time points. **(B)** agarose gel analysis for PCR-amplified ChIP-exo-TT libraries using U2OS whole cell extract (WCE) and nuclear isolate (NI) for immunoprecipitation with different antibodies. Input chromatin-DNA was processed similar to IP samples, but without exonuclease treatment (No-exo). **(C)** agarose gel analysis for H3K9me2 libraries of WT, RNase H overexpression (RNH OE) and WT U2OS cells carrying an empty plasmid (empty-vector), showing strong amplification for low-input library of RNH OE cells (300 k cells).

Surprisingly, PCR amplification produced strong signal after H3K9me2 ChIP-exo-TT for this very low cell number (Figure 3.8C). Notably, a slight difference in the size of PCR bands was observed for the different samples (WT, RNH OE and empty vector). This is due to the difference in the concentration of input DNA used, which led to difference in amplification, i.e., the size of PCR band for amplified ChIP library depends on the PCR cycle number and concentration of input DNA. Poor DNA or weak amplification produces lower bands as in

RNH OE low-input sample. This explanation is also supported by the fact that RNH OE sample shows strong primer dimers confirming that it has a lower amount of DNA. A strong evidence for this explantion is clearly seen in Figure 3.12A where the same library and the same input DNA produced different-size bands based on PCR cycle number. Also, all figures in this chapter show that poor ChIP libraries produce lower PCR bands compared to rich ones.

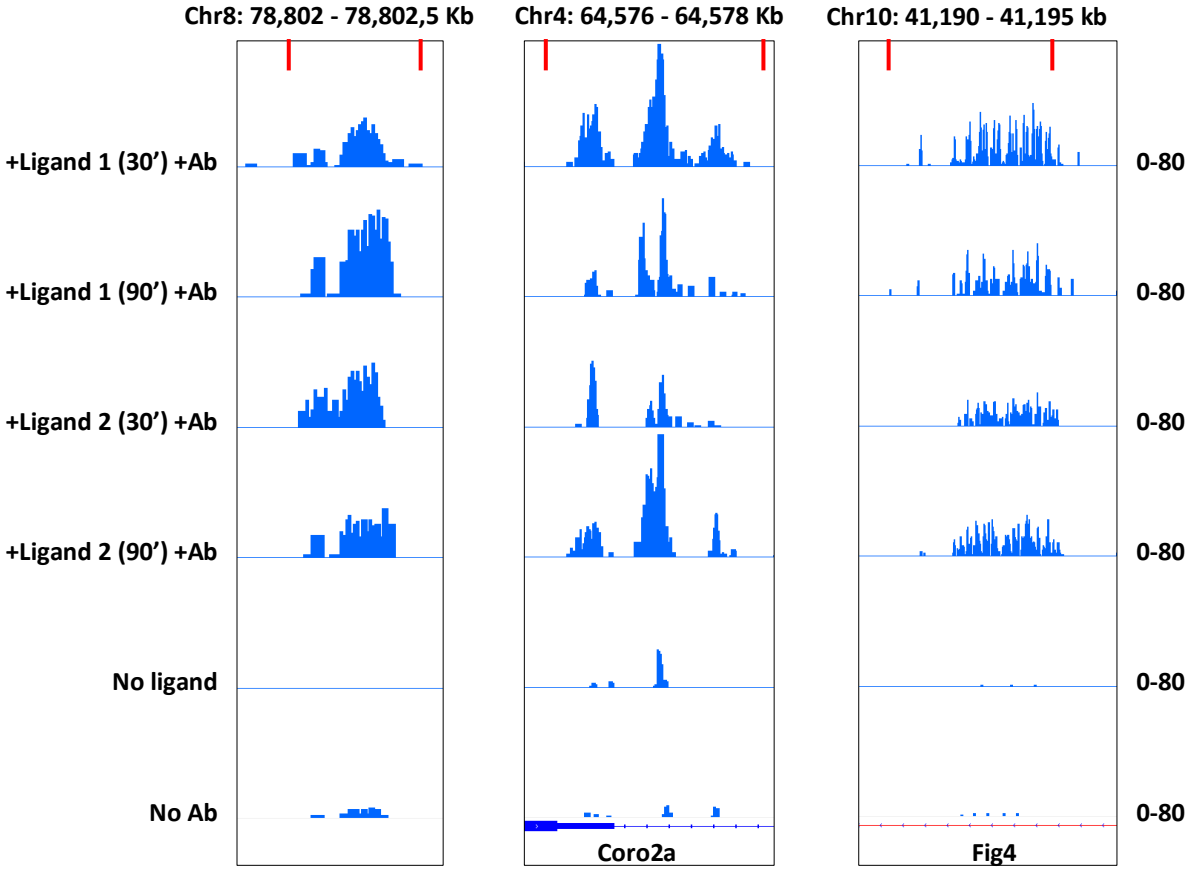



**Figure 3.9. ChIP-exo-TT captured ligand-activated AhR4 signals in mouse El4 lymphocytes**. Snap shot of genome browser showing signals enriched over genes know to be expressed in the same cell line. The signals weren't detected for the No-Ab and the no-ligand samples. EL4 cells were treated in four groups with one of two different ligands, each for two different incuation periods (30' and 90').

After sequencing, I used the data of this experiment to check H3K9me2 signals which were successfully detected for all cell types (Chapter 5). Based on these results, I deployed this method for a collaborative project aiming at mapping Aryl hydrocarbon Receptor (AhR) in mouse EL4 lymphocytes. AhR is a ligand-activated transcription factor involved in multiple biological processes including lymphocyte differentiation **(Larigot *et al.*, 2018)**. EL4 cells were treated with one of two different ligands (confidential information), each for two different periods (30' and 90') in order to activate AhR and induce its potential DNA binding for transcription activation. Interestingly, ChIP-exo-TT showed clear signals at specific genomic

regions for all ligand-treated cells, but not for the untreated or the no-antibody control (Figure 3.9).

### 3.2.8 SMART-seq enhanced the sensitivity of ChIP-exo (ChIP-exo-SMART) and abruptly shortened the time

The SMART technique (Switch Mechanism at 5' ends of RNA Transcript) is a rapid and high sensitivity library preparation method that was initially developed for RNA sequencing **(Y. Y. Zhu *et al.*, 2001; Zhumabayeva *et al.*, 2001)**. Recently this technique has been used to sequence DNA **(Turchinovich *et al.*, 2014; Vardi *et al.*, 2017)** after adding a poly-dA tail to the 3' end of DNA which I'm applying here. This technique relies on the use of a reverse transcriptase which has the ability to add non-templated deoxy-cytosines (dCs) to the 3' end of the cDNA. These dCs are required for the template switching through the annealing of template switching primer (TSP) which has three riboguanosines (rG)3 for the backward extension by the reverse transcriptase in the same reaction mix. Beside the fact that SMART technology is used for single cell RNA-seq assays due to its great sensitivity, it's also a single-tube compatible library preparation method. This means no purification steps are required in between, which would decrease loss of target DNA, enhance sensitivity and shorten the time. For these reasons, I sought to apply the SMART technology for ChIP-exo workflow. I called the new ChIP-exo version with SMART as ChIP-exo-SMART (Figure 3.10).

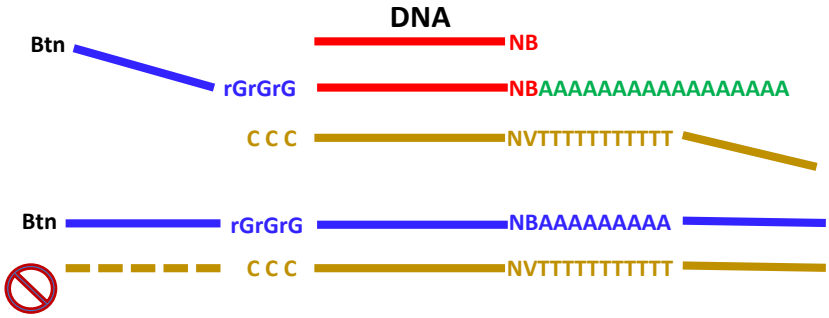



**Figure 3.10. ChIP-exo-SMART work flow.** After poly-dA tailing for DNA using Terminal deoxy-Transferase, a reverse transciptase is used to synthesize 1st strand cDNA and add 3Cs to its 3' for template switching using a primer that contain 3 rGs. Template switching primer is biotinylated (Btn) to block further reverse transcription and template switching rounds at the 3' end of 1st strand cDNA.

To develop and optimize ChIP-exo-SMART, I used fission yeast cells and different antibodies, H3K36me3 and H3K9me2, for the immunoprecipitation. I carried out all steps similar to ChIP-exo-TT with T7 exonuclease as the only on-bead treatment. After elution, I

carried out poly-dA tailing, reverse transcription (RT) and templates switching (TS) in a single tube without any purification steps. Surprisingly, the PCR amplification produced very strong signals after ChIP-exo-SMART for H3K36me3 DNA. However, no PCR signal was detected for H3K9me2 (Figure 3.11A). Since I was able to capture and map H3K9me2 with ChIP-exo-TT which has the same steps including the dA-tailing, I speculated that there may be a problem with either the reverse transcription or template switching. An important factor that can affect the RT step is contamination from the dA-tailing which can inhibit the reverse transcriptase. Independently, the TS step may require specific DNA end modification.

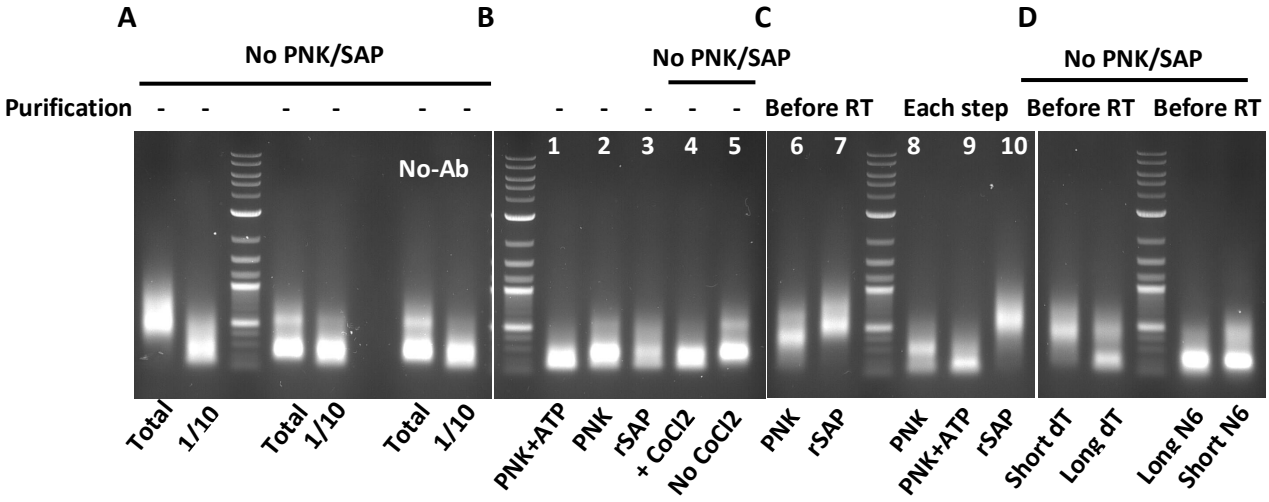



**Figure 3.11. ChIP-exo-SMART has superior sensitivity that can be even enhanced by DNA dephosphorylation and purification before RT**. Agarose gel analyses for ChIP-exo-SMART PCR-amplified libraries. All condition used for lanes 1 – 10 are listed in Table 3.5. Optimization was done with H3K36me3 antibody. Check signals for rSAP-treated and purified samples against the rest of conditions.

To test these possibilities, I decided to further optimize the SMART using one tenth of regular DNA amount from H3K36me3 IP. After elution, I treated the DNA with either rSAP or PNK to dephosphorylate both ends or phosphorylate 5'end of DNA, respectively. Separately, I conducted the dA-tailing without CoCl2 salt in tailing reaction buffer. Interestingly, I noticed a better PCR signal for the no-CoCl2 sample (Figure 3.11B), referring to the fact that a purification step is required before the RT step. To prove this, I repeated the above steps, but applied DNA purification at different intervals (before the RT, or before DNA modification step and RT). Strikingly, strong PCR signals were produced for rSAP-treated and pre-RT purified DNA (1/10 usual DNA amount) (Figure 3.11C). This strongly proves that DNA purification enhances the RT and that DNA has to be dephosphorylated. Using the same conditions, I found that short oligo-dT extension primer produces better results compared to

long one. Finally, I tried random reverse transcription using primer with six random nucleotides for extension, which was less efficient that oligo-dT primer (Figure 3.11D).

**Table 3.3. All conditions and steps used to optimize the SMART step after ChIP-exo**. This table lists conditions for lanes 1 – 10 in agarose gel analyses shown in Figure 3.11. PNK or rSAP were carried out at 37 °C / 20 min - 65 °C / 20 min while tailing reaction was carried out at 37 °C / 30 min - 70 °C / 10 min.

| Reaction / Sample | rSAP, PNK or PNK + ATP | Purif. | Poly-dA tailing | Purif. | SMART |
|---|---|---|---|---|---|
| **To test single-tube competency and check DNA end requirements** | | | | | |
| **1** | DNA, 1X **FS** buffer, 1 mM **ATP**, 10 U **PNK**, (10 µl total). | **No purification** | Modified DNA, 1X **FS** buffer, 1X CoCl2, 100 µM dATPs, 20 U TdT, (15 µl total). | **No purification** | Tailed DNA, 1 µl 10 mM dNTPs, 1 µl 5 µM short Uni.PE.dT24VN (17 µl total volume) Annealed at 95 °C / 3 min - 50°C / 3 min, snap chilled on ice then added; 1.5 µl 5X FS buffer, 2 µl 100 mM DTT, 100 U SSII, (21 µl volume). Run at 42 °C for 15 min then brought to room temperature before adding 1 µl of 10 µM TSP and incubating at 42 °C for one hour and inactivating at 70 °C for 15 min |
| **2** | DNA, 1X **FS** buffer, 10 U **PNK**, 10 µl volume | | | | |
| **3** | DNA, 1X **FS** buffer, 1 U **rSAP**, 10 µl volume | | | | |
| **To test inhibition by CoCl2 (presence vs absence)** | | | | | |
| 4 | **No treatment** | **No purification** | DNA, 1X **FS** buffer, 1X **CoCl2**, 100 µM dATPs, 20 U TdT, (15 µl total). | | |
| 5 | **No treatment** | | DNA, 1X **FS** buffer, **no CoCl2**, 100 µM dATPs, 20 U TdT, (15 µl total). | | |
| **To investigate DNA end requirements and test purification requirement before RT** | | | | | |
| **6** | DNA, 1X **TdT buffer**, 10 U **PNK**, 10 µl volume | **No purification** | Modified DNA, 1X **TdT** buffer, 1X CoCl2, 100 µM dATPs, 20 U TdT, (**25 µl total**). | **AMPure XP beads purification, 1.8 volumes, elution in 14 µl** | Tailed DNA, 1 µl 10 mM dNTPs, 1 µl 5 µM short Uni.PE.dT24VN (16 µl total volume) Annealed at 95 °C / 3 min - 50°C / 3 min, snap chilled on ice then added; 5 µl 5X FS buffer, 2 µl 100 mM DTT, 100 U SSII, (24 µl volume). Run at 42 °C for 15 min then brought to room temperature before adding 1 µl of 10 µM TSP and incubating at 42 °C for one hour and inactivating at 70 °C for 15 min. |
| **7** | DNA, 1X **TdT buffer**, 1 U **rSAP**, 10 µl volume | | | | |
| **To investigate DNA end and purification requirements** | | | | | |
| **8** | DNA, 1X **PNK** buffer, 1 U **PNK**, 20 µl volume | **AMPure XP beads purification, 1.8 volumes, elution in 14 µl** | Modified DNA, 1X **TdT** buffer, 1X CoCl2, 100 µM dATPs, 20 U TdT, (**25 µl total**). | | |
| **9** | DNA, 1X **T4 DNA ligase** buffer, 10 U **PNK**, (20 µl total). | | | | |
| **10** | DNA, 1X **Cutsmart** buffer, 10 U **rSAP**, 20 µl volume | | | | |

### 3.2.9 ChIP-exo-SMART overamplifies background noise which masks target signal

These results were tempting to try this method in a collaborative work to map different FLAG-tagged proteins in HEK293 mammalian cells. Unexpectedly, PCR amplification produced very strong signals for no-antibody library. This signal was reduced but not eliminated with lower PCR cycles (Figure 3.12A). The difference in the size of bands produced using different PCR cycles for the same samples using the same DNA amount cofirms my explanation for a similar observation in Figure 3.8 for amplification of WT, RNH OH and empty vector ChIP-exo-TT samples. Generated NGS sequencing data revealed enhanced noise masking positive signals. However, qPCR analysis using eluted DNA from the same samples, showed expected enrichment over target sequences (data not shown). I encountered the background noise problem when I tried to map AhR in EL4 mouse cells with ChIP-exo-SMART instead of ChIP-exo-TT which previously showed sepcific signals in positive experimental samples, but not in negative-control samples (Sectiton 3.2.7).

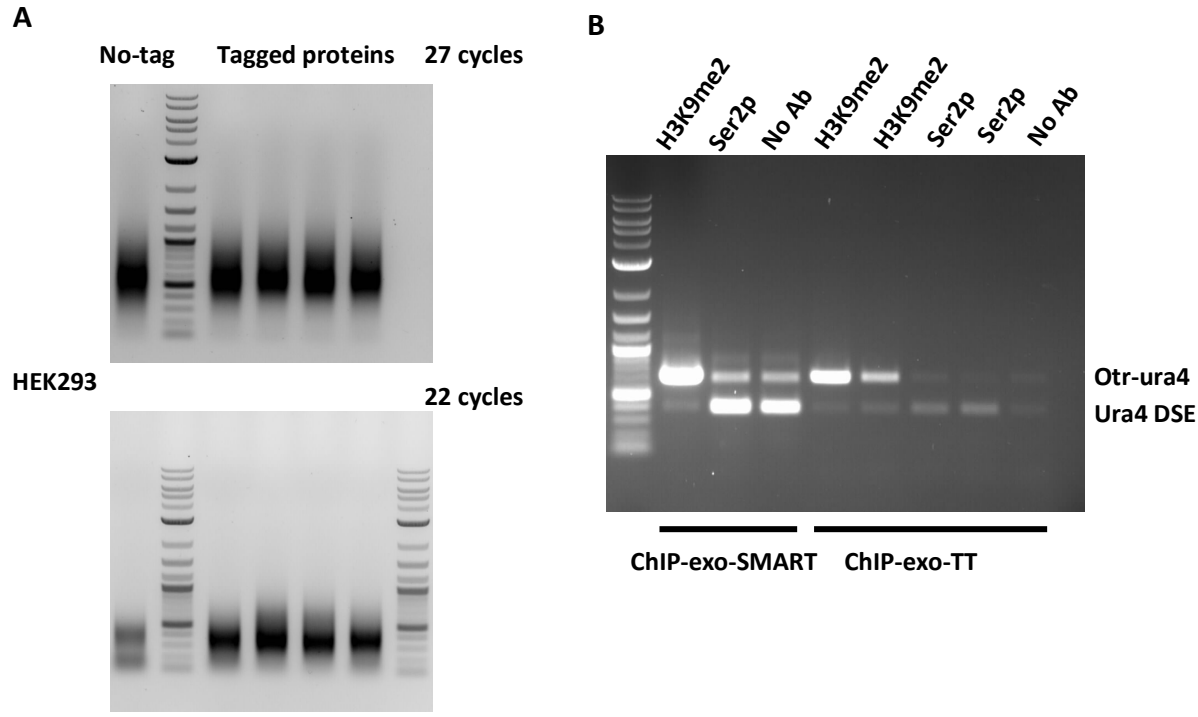



**Figure 3.12. ChIP-exo-SMART induces DNA overamplification and enhances background signal. (A)** PCR amplification produces strong signal in no-tag negative control compared to positive libraries. **(B)** PCR amplification for ChIP-exo-TT libraries produced single band specific for one ura4 locus as expected with each antibody. However, two PCR bands specific for both loci are produced for ChIP-exo-SMART libraries. ChIP-exo-TT No-Ab library is empty compared to that of ChIP-exo-SMART which shows strong amplification.

Furthermore, I tried to map H3K9me2 mark and Ser2-phosphorylated Pol II in fission yeast using both ChIP-exo-TT and ChIP-exo-SMART methods. At the end of the experiments, I used the final DNA library to validate the results using standard PCR in order to test enrichment of these factors at heterochromatic otr-ura4 and euchromatic ura4 DSE. A PCR signal should be produced for otr-ura4 (at heterochromatin) with H3K9me2 but not Ser2p library while the opposite is true for ura4 DSE. Indeed, this was confirmed for ChIP-exo-TT libraries. However, ChIP-exo-SMART libraries showed enhanced background signals for all samples including no-antibody control which was nearly empty in ChIP-exo-TT (Figure 3.6 & Figure 3.12). These observations show that the SMART enhances background signal. It can't be excluded that the SMART may represent an early amplification step before the final PCR. For this reason, I stopped ChIP-exo-SMART optimization and decided to depend on ChIP-exo-TT for mapping different proteins.

## 3.3 Discussion

### 3.3.1 ChIP-exo versions

Because of the relatively high background contamination from protein-unbound DNA, classic ChIP may produce false-positive peaks (false positives or erroneous calls) which may sometimes require stringent filtering that may otherwise lead to loss of real, but weak signals (false negatives or missed positive calls). ChIP-exo technique significantly diminishes these pitfalls by providing a near single-base pair resolution with minimal noise to signal ratio **(Rhee & Pugh, 2011**, **2012)**. Genuinely, I wanted to study the global enrichment of H3K9me2 histone mark in fission yeast in response to changes in R-loop levels, specifically, at gene terminators and in relation to transcription termination. In yeast, gene terminator regions are very short, and transcription termination window is very narrow compared to those of mammalian cells. This necessitated the use of a high-resolution method for precise mapping. For this reason, I sought to use ChIP-exo technique instead of classic ChIP.

However, original ChIP-exo work flow involves around 13 reactions that can extend for a whole week, making it a long and very laborious technique. ChIP-exo reactions are split into two stages, on-bead reactions starting just after immunoprecipitation till elution step, and in-solution reactions starting after DNA extraction till library preparation. On-bead reactions are intervened by multiple rotation and washing steps, while in-solution reactions are separated by

AMPure XP beads purification steps. These frequent washings and purifications may affect the sensitivity as it may lead to loss of some DNA material and depletion of weak signals.

It's noteworthy that another version called ChIP-nexus was developed to improve these aspects **(He *et al.*, 2015)**. For ChIP-nexus, after the first adaptor ligation and lambda exonuclease treatment, the ssDNA is circularized through intramolecular single stranded ligation. Next, a primer is annealed to the ligated adaptor sequence in circular ssDNA to make dsDNA. The circular DNA is then digested at the generated dsDNA sequence of the adaptor. The linearized ssDNA will have two adaptor sequences at both ends and ready for direct amplification and sequencing **(He *et al.*, 2015)**. Compared to original ChIP-exo, ChIP-nexus didn't show a significant improvement, but suffered from loss of data. This loss has been interpreted as a result of incomplete synthesis of barcodes sequence during oligo manufacturing process as the barcodes were designed at the 3' end of adaptors **(Rossi *et al.*, 2018)**.

Recently, different simplified versions of ChIP-exo have been introduced to improve different aspects including sensitivity, low input-competency, and time-efficiency while preserving the high resolution and low background features of the original method **(Rossi *et al.*, 2018)**. The hallmark of the new versions was the replacement of the double stranded A-T ligations with single-stranded splint ligation steps. The main difference between the different versions of ChIP-exo is the types of ligations used. While two splint ligations are used to ligate splint adaptors to ssDNA as in ChIP-exo-(SL)2, a mix between these two forms is used in ChIP-exo-SL-DS (first splint and second A-T ligation) and ChIP-exo-DS-SL (first A-T, and second splint ligation).

### 3.3.2 Ligation conditions and requirements

#### *3.3.2.1 A-T ligations*

Notably, double-stranded A-T ligation requires more steps than splint ligations. For A-T ligation, DNA ends have to be blunt, 5' ends of either the DNA or the adaptors have to be phosphorylated, and 3' ends of the DNA have to be dA-tailed to stick to the dT overhang of the adaptors. These conditions are required for the success of the A-T ligation. As a prerequisite, on-bead polishing and kinase reactions must be done at the beginning. These can be combined into one step similar to illumina DNA library preparation workflow which I applied for an improved ChIP-exo version. I found that Combining these steps greatly increased the sensitivity and improved the efficiency of the experiment (Figure 3.4C).

Similarly, the second A-T ligation requires blunt, phosphorylated, and dA-tailed dsDNA. Since the eluted DNA is mainly single stranded, but partially double stranded at 5' ends, a fully double stranded DNA has to be generated by primer extension. For primer extension, I tried two different enzymes, Bst and Phi29 DNA Polymerase, separately. I found that extension with Phi29 DNA Polymerase followed by dA-tailing by Klenow exo- DNA Polymerase produced better results compared to extension and dA tailing by Bst in a single reaction. This may be due to a difference in the activity of enzymes and efficiency of each step. It may be also due to a difference in reaction temperature which is high in case of Bst. This high temperatue may decrease the annealing of poly-dTVN extension primer which has low melting temperature.

#### *3.3.2.2 Splint ligations*

On the other hand, splint ligation doesn't impose specific requirements except the presence of ssDNA. Consequently, this eliminates the critical need for blunting, dA-tailing and phosphorylation. Importantly, this shortens the time and omits many washing and purification steps that may deplete DNA. However, I found ChIP-exo-(SL)2, which has the simplest and shortest workflow, to be less robust than the original ChIP-exo. Strikingly, splint ligation produced substantially strong adaptor dimers that may only be removed by gel purification (Figure 3.4A, B). A clear observation evidencing for the low efficiency of splint ligation is that the use of one A-T and one splint ligation increased the yield and eliminated the dimers (Figure 3.4C). Obviously, the success of the whole experiment seems to be more dependent on the efficiency of the first ligation. This was confirmed by the observation that incorporating the double stranded ligation before the splint ligation greatly reduced the adaptor dimers and improved the results compared to performing these ligations in the reverse order (Figure 3.4).

Although I haven't tested the reasons for the low competency of the splint ligations, my estimation is that the used conditions aren't optimal enough. Non-optimal conditions may include the DNA ligase, structure of adaptors, incubation time or other factors. T4 DNA ligase may not be the favoured enzyme for ligating this kind of DNA ends. Regarding the splint adaptors, an overhang of 5 random-nucleotides may not be optimal for the splint ligation. At least six nucleotides are required for the success of splint ligation **(Agarwal *et al.*, 2015; Gansauge *et al.*, 2020; Raine *et al.*, 2017)**. It's worth mentioning that different from simplified ChIP-exo versions, which didn't use blocked adaptors, I used adaptors with blocked 3' ends for all ligations. This condition is essential for suppressing the formation of adaptor dimers as

recommended in many protocols **(Agarwal *et al.*, 2015; Gansauge *et al.*, 2020; Raine *et al.*, 2017)**. Even with the use of blocked adaptors, splint ligation still forms strong empty adaptor dimers (Figure 3.4).

It has been shown that splint ligation suffers from severe sequence bias, low efficiency and poor yield **(Kwok *et al.*, 2013)**. Beside the fact that adaptor ligation efficiency is known to be biased to the sequence of ligated nucleic acid ends, different factors may affect the efficiency of ligation in general and splint ligation in particular. The efficiency of splint ligation may differ according to the length of the overhang and, more importantly, according to the direction of the reaction (whether the overhang exists on the 3' or the 5' end of the DNA). For these doubts, I sought to replace ligation steps with ligation-free approaches.

### 3.3.3 Exonucleases

For the exonuclease treatment, the brand-name reaction step for the ChIP-exo, lambda exonuclease is commonly used. The exonuclease step comes after the first ligation in original ChIP-exo and ChIP-exo-DS-SL, and before both ligation steps in ChIP-exo-SL2 and ChIP-exo-SL-DS. It also comes before poly-dA tailing in ChIP-exoTT and ChIP-exo-SMART. Because blunt and phosphorylated DNA ends are required for the action of the lambda exonuclease, these steps are still included in splint ligation and ligation-free ChIP-exo versions. An alternative for lambda exonuclease is T7 exonuclease which has the advantage of not requiring neither double stranded nor phosphorylated ends. T7 exo is known to have less processivity compared to lambda exo, yet I have found that it maintains the single nucleotide resolution of ChIP-exo. This has been revealed in budding yeast Kre33 mapping signals (Figure 3.7), and fission yeast and mammalian cell H3K9me2 mapping signals (Chapter 5).

Another exonuclease is the Rec$_{Jf}$ that degrades single stranded DNA. It has been reported that this step can be excluded as it doesn't make a difference **(Rossi *et al.*, 2018)**. This is consistent with my results for R-loop mapping for crosslinked and non-crosslinked samples where I, respectively, included or excluded this step. There was no remarkable difference in R-loop mapping profiles for both conditions. Excluding Rec$_{Jf}$ step, especially for the A-T ligation versions, shouldn't not present any problem as A-T ligations will have minimal efficiency for this kind of template.

### 3.3.4 Poly-dA tailing, a sensitive ligation-free approach for ChIP-exo-TT method

One strategy that represents a great advance over ligation is poly-dA tailing which seems to have higher efficiency and less sequence bias compared to ligation. In addition, tailing doesn't have prerequisites as it can be done for dsDNA or ssDNA template and doesn't require long times compared to ligation. For ChIP-exo-TT, on-bead exonuclease step is done before poly-dA tailing. After elution and poly-dA tailing, extension is done using oligo-dT primer to generate dsDNA for the rest of steps to be done similar to original ChIP-exo till A-T ligation.

ChIP-exo-TT proved to have similar high resolution, but higher sensitivity compared to the original and all ligation-based methods. Multiple facts such as absence of background noise and fragment size distribution for amplified libraries (bioanalyzer profiles) support the high resolution which is clearly seen in the genome browser signals. R-loop profiling studies (Chapter 4) strongly evidences for the high resolution of ChIP-exo-TT compared to original ChIP-exo. Only, using ChIP-exo-TT, could I map H3K9me2 in fission yeast (Figure 3.6 and Chapter 5), Kre33-FTP in budding yeast (Figure 3.7) and AhR in El4 lymphocytes (Figure 3.9). A great benefit for the ChIP-exo-TT is that it can produce more directional signals (results from R-loop mapping in Chapter 4) which means it can reveal ssDNA-binding bias for different DNA-binding proteins which may bind to one DNA strand. The fact that ChIP-exo-TT doesn't require dsDNA, makes it more convenient for mapping ssDNA-binding proteins.

One disadvantage for the ChIP-exo-TT is that poly-A tracts may be erroneously enriched as a result of sequencing the dA tail of the extension primer. Trimming the A-tail may trim specific A-rich sequences. A possible solution is to use a sequencing primer that binds directly to the A-tail just before the enriched DNA sequence. In principle, ligations can be skipped and two poly-dA tailing steps may be used to add different tails instead **(D. Jiang *et al.*, 2019)**, however this may require special considerations to make sure this will not perturb clustering or eliminate strand specificity.

### 3.3.5 SMART technology for a totally ligation-free ChIP-exo-SMART

With these motivating results, I tried to further improve ChIP-exo-TT and replace the final ligation step with another ligation-free method. For this, I used the SMART technology, a ligation-free method widely used for its ultra-high sensitivity and superior performance enabling sequencing various RNA and DNA templates **(Fish *et al.*, 2016; Hagemann-Jensen**

***et al.*, 2020; Picelli *et al.*, 2013; Picelli *et al.*, 2014; Vardi *et al.*, 2017; Verboom *et al.*, 2019; Zhao & Zheng, 2018; Y. Y. Zhu *et al.*, 2001; Zhumabayeva *et al.*, 2001)**.

ChIP-exo-SMART starts similar to ChIP-exo-TT, but differs in the reverse transcription and template switching steps which come after tailing (Figures 3.5 & 3.10). Surprisingly, the SMART proved to be compatible with single-tube workflow. However, some modifications were required for low-input samples; the DNA had to be dephosphorylated and purified to enhance reverse transcription and template switching (Figure 3.11 and Table 3.3). ChIP-exo-SMART improved the amplification and DNA size distribution compared to the ChIP-exo-TT method. Unfortunately, it also enhanced the amplification and signal of the no-antibody control which never gets clean. The SMART reaction may present an unwanted pre-PCR amplification step. It's not clear how this can happen, especially, if the template switching primers are biotinylated to block further reverse transcription and template transcription rounds at the 3' end of first cDNA strand (Figure 3.10). Since the first steps of ChIP-exo-SMART are done similar to the ChIP-exo-TT, I expect the RT and TS steps to be the reason for these issues.

A problem with the SMART method is that the generated reads may have 5' and 3' ends with low complexity due to the presence of 3G repeats and poly-dT tails, respectively, which may affect the clustering. For this reason, an external spike-in DNA may need to be added to increase the complexity of the library. A possibility that should be considered, and may need to be tested, is that RT enzyme may fall before generating the full-length cDNA, and produce incomplete cDNA copies (truncated reads) that may affect the mapping. In spite of all the above doubts for DNA-based SMART, the same technique worked efficiently for RNA sequencing. Surprisingly, I used it for mapping RNA of R-loops with high sensitivity, high resolution and specificity (Chapter 4). An important fact is that the final SMART R-loop libraries were pooled with other ChIP-exo-TT libraries before sequencing which may have increased the complexity and enhanced the clustering for SMART reads.

In comparison to ChIP-exo, Tn5 transposase was used to cleave ChIPed DNA (no lambda exo) and directly ligate adaptors to it. Although this strategy showed higher sensitivity, it was reported to compromise the resolution **(Rossi *et al.*, 2018)**. Other techniques such the nuclease-targeted CUT&RUN method **(Skene & Henikoff, 2017)** may represent high sensitivity and time-saving methods but may not have the same resolution as ChIP-exo. My work, similar to

others, opens the gate for harnessing different approaches for high-sensitivity and high-resolution mapping of protein-DNA complexes at genome-wide and site-specific levels.

## 3.4 Conclusion

With results obtained by ChIP-exo-TT, I can confidently say that it's superior to all previous ChIP-exo methods in sensitivity. I used ChIP-exo-TT to map different marks and proteins such as modified histones (H3K9me2) and transcription factors (Kre33 and AhR). Especially, H3K9me2 couldn't be captured by any of previous methods. Excitingly, ChIP-exo-TT successfully enriched Kre33, a single-locus DNA-binding protein, and detected subtle differences in its binding to the ETS, a very short region of rDNA, in WT and mutant budding yeast which was confirmed by qPCR. It also showed high performance with low-input samples.

ChIP-exo-TT isn't a completely new technique, it's a modification of the ChIP-exo method; it follows the same principle and steps including cell lysis, sonication for chromatin, polishing and kinase and more importantly, the lambda exonuclease treatment which is required for trimming the 5' end for high resolution. ChIP-exo-TT also preserves the same steps (extension, and dA-tailing) toward dsDNA adaptor ligation to the 5' end of trimmed DNA, so it just introduces poly-dA tailing as a different approach for capturing 3' end of DNA. This is a very important point as the 5' end, but not 3' end is used for read aligning. It's important to note that the authors of ChIP-exo compared the resolution of all versions and found that they all have high resolution as long as the exonuclease treatment is included. Although they reported a slight difference based on whether exonuclease was done before or after first ligation, they concluded that all versions have the same lambda exonuclease pattern. Resolution was compromised only when Tn5 transponsase without lambda exo was used for tagmentation/ChIPmentation **(Rossi *et al.*, 2018)**. This makes me sure that ChIP-exo-TT has high resolution. This can be confirmed by inspecting signal ends, the very low noise, and even in the bioanalyzer profiles (Chapter 5) which exhibit smaller average size (200 bp) compared to original ChIP-exo (300) which suggests that ChIP-exo-TT may even have higher resolution.

Finally, further investigation is required to fix issues such as low complexity and high background of the DNA-based ChIP-exo-SMART. This makes ChIP-exo-TT the method of choice for DNA-based ChIP-exo applications at the current moment.

# Chapter 4. Investigating Putative Roles for R-Loops in dsRNA Formation and Gene Regulation Using New High-resolution and Directional Mapping Techniques

## 4.1 Introduction: R-loop profiling and antisense transcription

Under specific genetic deficiencies and pathological conditions, R-loops accumulate in the nucleus to increase recombination and mutation rates, and induce DNA damage and genomic instability. For this reason, R-loops have been considered for long as transcriptional byproducts that threaten genomic integrity **(Aguilera & Garcia-Muse, 2012)**. However, R-loops form also in wild-type cells (under physiological conditions), but at lower levels compared to mutant cells. R-loops have been reported to form over specific genomic regions such as highly transcribed ORFs, retrotransposons, tDNA arrays, small nuclear and nucleolar RNA genes, ribosomal DNA repeats, telomeric repeats, centromeres and mitochondrial genes **(Garcia-Muse & Aguilera, 2019; Santos-Pereira & Aguilera, 2015)** (Figure 4.1). Moreover, they correlate with different epigenetic marks in yeast **(Hartono *et al.*, 2018)**, mammals **(L. Chen *et al.*, 2017; Sanz *et al.*, 2016)** and plant cells **(Xu *et al.*, 2017)** implying that R-loops are programmed and scheduled rather than arbitrarily-formed structures. Intriguingly, they have been also mapped to specific genic compartments among which are non-methylated CGI promoters **(Ginno *et al.*, 2013; Ginno *et al.*, 2012)** and specific terminator regions in

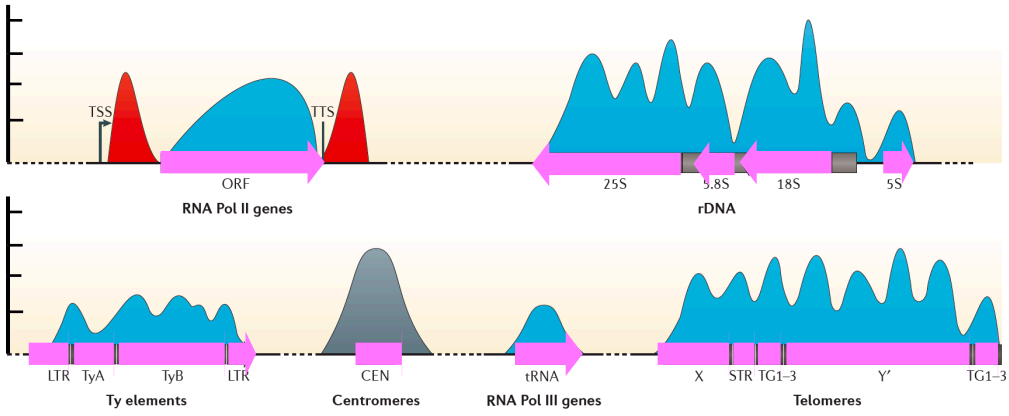



**Figure 4.1. Hot spots for R-loop formation in yeast and mammalian cells.** Schematic representation for R-loops profiles over highly enriched regions. In mammalian cells, R-loops form over gene promoters, gene bodies of a subset of genes and GC rich terminators. In yeast, R-loops accumulate over rRNA and tRNA genes, transposon elements, telomeres, and centromeres. Adapted from **Santos-Pereira and Aguilera (2015).**

mammalian cells **(Sanz *et al.*, 2016; Skourti-Stathaki *et al.*, 2014)** suggesting that R-loops have different transcriptional regulatory roles at these different genic regions.

Of spectacular interest is the finding that overexpression of RNase H1 which degrades R-loops, eliminates dsRNA and depletes RNA interference machinery factors such as Dicer 1 (Dcr1), Argonaute protein 1 (Ago1), G9a histone lysine methyltransferase, H3K9me2 histone mark, and the silencing heterochromatin protein 1γ (HP1γ) over mouse β-Actin gene terminator **(Skourti-Stathaki *et al.*, 2014)**. These findings suggested that R-loops induce antisense transcription over gene terminators and lead to dsRNA formation which recruits the RNA interference (RNAi) machinery for heterochromatin deposition and efficient termination **(Skourti-Stathaki & Proudfoot, 2014)**. However, such a role and a correlation between R-loops and dsRNA haven't been tested on a genome-wide scale. Nevertheless, the exact mechanism by which antisense transcription is induced and dsRNA is formed, is still ambiguous (Figure 4.2).

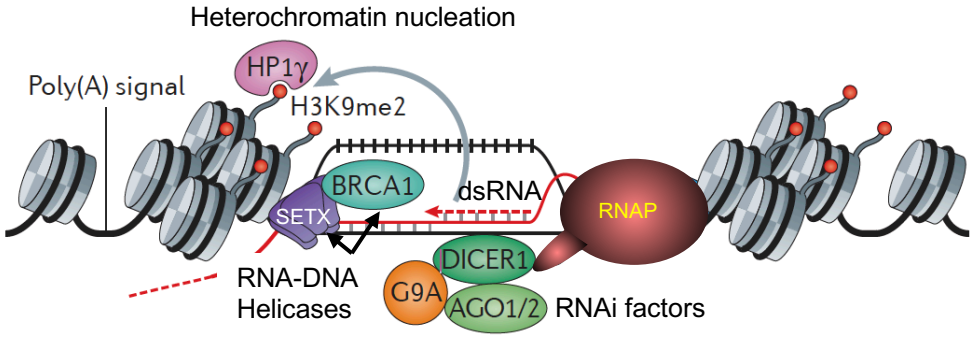



**Figure 4.2. A proposed model for R-loop-mediated dsRNA formation, heterochromatin nucleation and transcription termination.** According to this model, formation of an R-loop downstream poly (A) signal induces antisense transcription and double stranded RNA formation. This dsRNA mediates heterochromatin nucleation by recruiting the RNA interference factors such as Dicer 1, Argonaut protein 1 (Ago1) and histone methyltransferase G9a which deposits H3K9me2 histone mark and triggers the loading of heterochromatin protein 1γ (HP1γ) required for termination. Different factors such as BRCA1 and SETX are suggested to participate in this termination process in mammalian cells. Adapted from **Santos-Pereira and Aguilera (2015).**

Intriguingly, the genome of any living organism is pervaded by different forms and types of noncoding RNA and cryptic transcripts **(Jensen *et al.*, 2013)** that have been recently revealed using advanced sequencing technologies (Figure 4.3). Although they have been considered for long as junk transcripts, the functions of some of these non-coding transcripts have recently been deciphered **(Kopp & Mendell, 2018; Ransohoff *et al.*, 2018; Statello *et al.*, 2021; Till *et al.*, 2018)**. In yeast, R-loops have been found to be enriched at antisense transcription-associated genes whose expression was significantly altered upon RNase H

overexpression **(Chan *et al.*, 2014)**. In addition, a partial cause-effect correlation between R-loops and some antisense transcripts has been recently revealed by Proudfoot lab showing that R-loops promote antisense transcription over the genome of mammalian cells **(Tan-Wong *et al.*, 2019)**. It's tempting to think that some of these antisense transcripts may contribute to the formation of dsRNAs along with the RNA-DNA hybrid. Yet, it is not known whether any kind of interaction exists between RNA-DNA hybrids and the corresponding homologous antisense non-coding transcripts. Finally, whether an R-loop can initiate the formation of another antisense R-loop and consequently a double R-loop structure isn't known.

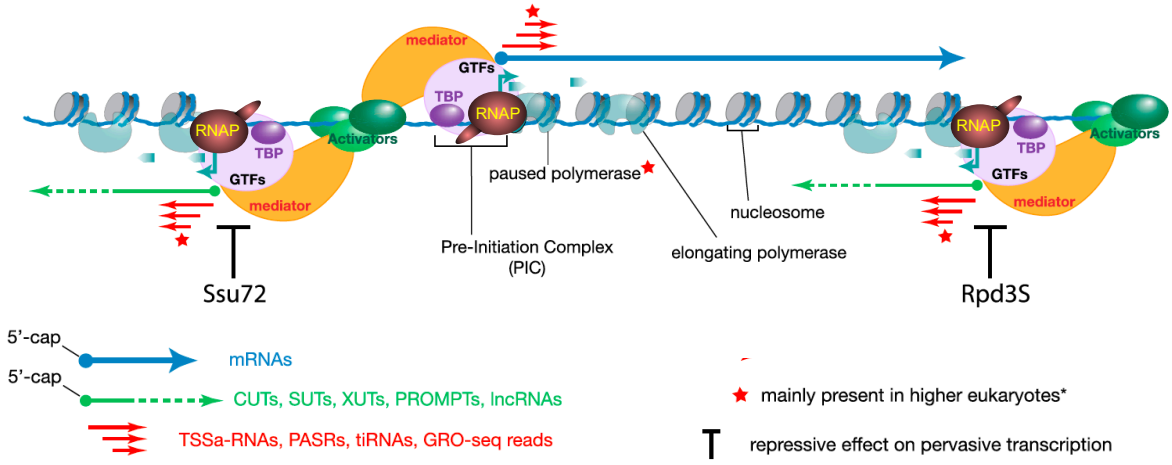



**Figure 4.3. Non-coding RNAs associated with mRNA transcription units.** GTFs, general transcription factors; TBP, TATA box-binding protein. Adapted from **Jensen *et al.* (2013).**

A major challenge against understanding possible correlations and associations between R-loops and antisense transcription, is the limitations of current R-loop mapping techniques and discrepancies among their results. Most of the current R-loop-mapping DRIP methods suffer from the lack of either resolution or directionality or both, except for the DRIPc and RR-ChIP methods which, regardless, may suffer from poor sensitivity and underrepresentation of signals as a result of depending on RNA for library preparation.

Advances in R-loop detection technologies have enabled uncovering many aspects about R-loops. These include formation conditions, control factors, genomic distribution for different organisms, the different functional implications and pathological conditions. Despite this great deal of knowledge, there is currently much uncertainty about many findings even about those corresponding to the same organism. This uncertainty arises mainly due to clear discrepancies among results, which may be explained by the wide range of technical variations and inherent biases in the methodologies by which different data were generated. Most of these findings are

concluded from R-loop mapping data generated by DNA-RNA hybrid immunoprecipitation (DRIP) using S9.6 antibody followed by either next generation sequencing or microarray analysis **(Halasz *et al.*, 2017; Vanoosthuyse, 2018)**. These variations, thought to be the reason for the contradictory results, include the model organism and cell type, cell cycle and growth phase of the cells, fixation step inclusion or omission, lysis method (mechanical homogenization or enzymatic proteolysis), immunoprecipitation input (chromatin preparation or nucleic acids extract), total nucleic acid extraction method, fragmentation method (restriction digestion or sonication), different enzymatic treatments, library preparation template (RNA or DNA), and library preparation method and the analysis algorithm **(Halasz *et al.*, 2017; Vanoosthuyse, 2018)**.

A recent study has conducted a comprehensive analytical workflow to investigate the impact of different DRIP conditions on the final outcome. Among other findings, it concluded that the use of restriction enzymes for fragmentation of genomic material compromises the resolution and produces long fragments over ORFs of coding genes, especially over the first exon. This drawback misled to the conclusion that R-loops accumulate at the first exon of coding genes. Consequently, results about evolutionary conservation of R-loop-forming sequences, inferred from sequence homology over R-loops-forming first exons, may not be precise **(Halasz *et al.*, 2017)**. Another more striking example is the formation of R-loops over gene terminators; while DRIPc method enriched R-loops at terminator regions of mammalian cells **(Sanz *et al.*, 2016)**, R-ChIP techniques which map R-loops indirectly through ChIPing catalytically defective RNase H1, showed enrichment of R-loops only at promoters **(L. Chen *et al.*, 2017; Tan-Wong *et al.*, 2019)**. In fact, the number of positive loci mapped for a specific organism depends on the method used and also differs from an experiment to another **(Santos-Pereira & Aguilera, 2015)**. This explains why different approaches may exhibit a range of inconsistencies and a wide gap between results.

## 4.2 Research aims and proposed models

Keeping these limitations in mind and carefully considering the fact that different methods can produce inconsistent and, probably, misleading results, I sought to deliver a directional and high-resolution R-loop profiling using multiple methods, yet, with a similar principle. For this, I adopted the chromatin-based DRIP-like workflow and harnessed the high-resolution ChIP-exo method combined with either double-stranded DNA (dsChIP-exo) or

single-stranded DNA (ssChIP-exo) capturing approaches to map R-loops in the fission yeast *S. pombe*. Adopting a similar chromatin-based workflow, I used the DRIPc method with some modifications including the treatment with multiple enzymes and the incorporation of the SMART technology to present another high-resolution, high-sensitivity and directional mapping method that is also complementary to the ssChIP-exo method.

Using multiple methods with possibly different propensities to capture different R-loop-forming sequences would enable capturing the biggest fraction of the genome (the union of regions) that may form R-loops. This would provide a comprehensive and complementary mapping. At the same time, using directional DNA-based and RNA-based mapping approaches should provide signals that are complementary in spatial coincidence and in direction of mapped sequences.

The next aim was to investigate, but on a genome-wide scale, a previously suggested role for R-loops in dsRNA generation. dsRNAs have a great scientific importance as precursors for RNAi machinery required for gene silencing and heterochromatin formation. If R-loops were involved in dsRNA generation, this would mean that R-loops may have a role in endogenous small interference (siRNA) biogenesis and RNAi-mediated gene silencing pathway. Accordingly, it's intriguing to understand the mechanism by which R-loops may contribute to dsRNA formation. Only two pathways can be envisaged for R-loop-dependent dsRNA formation. Here, I proposed two models representing two slightly different mechanisms for R-loop-mediated dsRNA formation (Figure 4.4). According to the first model, an RNA-DNA hybrid may induce the formation of an overlapping antisense RNA-DNA hybrid to form a bidirectional hybrid in a structure that can be called a double R-loop. The second model suggests that an RNA-DNA hybrid coupled to and overlapping with free antisense RNA transcription may induce dsRNA formation. The two models suggest that the displaced ssDNA strand of R-loop can support antisense transcription, but they slightly diverge according to the hybridization nature of transcribed asRNA. According to the first model, the asRNA may co-transcriptionally hybridize to its template DNA strand and form an antisense hybrid while the second model suggests that the asRNA will get co-transcriptionally separated and freed from the DNA and get somehow stabilized locally to form dsRNA at a later stage (Figure 4.4).

To test these models, I used the generated R-loop mapping data from the three methods to test the correlation between R-loop signals forming on both DNA strands and examine the formation of double R-loops genome-wide. I also studied the stability of R-loops against RNase H by comparing R-loop profiles of WT and *rnh1Δ rnh2Δ* yeast cells. In addition, I examined the association between R-loops, antisense transcription and dsRNA formation.

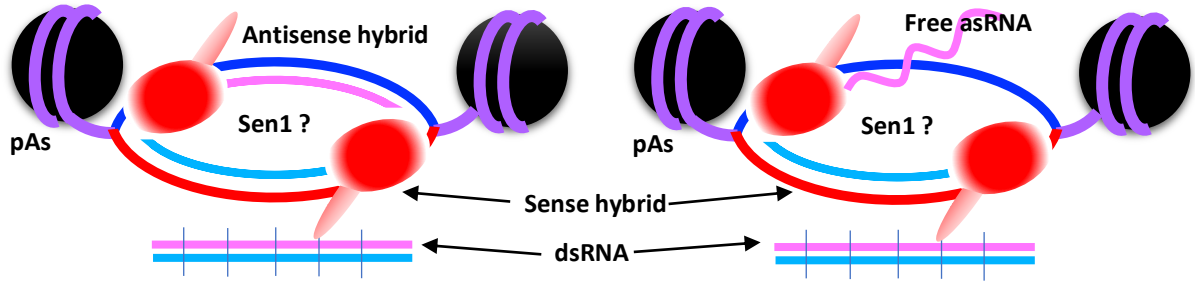



**Figure 4.4. Schematic representation for the two proposed R-loop-mediated dsRNA formation models**. Both models are based on RNA-DNA hybrid formation over one DNA strand and antisense transcription over the displaced free ssDNA. Left model suggests dsRNA formation through bidirectional RNA-DNA hybrid (double R-loops) formation while right model suggests dsRNA formation by single RNA-DNA hybrid formation and free asRNA. pAs, polyadenylation signal; Sen1, Sen1 helicase or a homologous protein.

Here, I show clear differences among the different R-loop mapping methods in directionality of detected signals. While dsChIP-exo suggested double R-loop formation genome wide, ssChIP-exo and SMART-DRIPc generated unidirectional signals that map mainly to template DNA strand, which doesn't support global double R-loop formation. Moreover, ssChIP-exo and SMART-DRIPc didn't support a correlation between forward- and reverse-strand R-loops. I also show that R-loops forming over 60 to 70 % of R-loop forming sequences are sensitive to RNase H in WT cells, not supporting dsRNA formation through these R-loops. However, some R-loop signals showed insensitivity to RNase H and formed over dsRNA-forming genes of sense antisense transcription. Unexpectedly, I found that RNase H deletion decreases R-loop signals over template strand of ribosome biogenesis and protein synthesis genes, suggesting an effect for RNase H manipulation on global transcription and gene expression. Moreover, I found that RNase H deletion affects R-loop signals on both DNA-strands in a contrasting way; while it increases R-loop signals over forward strand, it decreases signals over the reverse strand. Furthermore, I noticed a bias in sensitivity of R-loops to wild type levels of RNase H, based on orientation of R-loops (formation direction). Consistent with this, RNase H-insensitive R-loops showed a bias to form on reverse DNA strand. These results suggest a role for RNase H in regulating sense-antisense transcription. Excitingly, my

observations suggest a role for RNase H in regulating R-loops levels in a way dependent on R-loop orientation. Independently, my results point toward a role for RNase H in regulating gene expression based on transcription direction of affected genes.

## 4.3 Results

### 4.3.1 ssDNA can be transcribed to form RNA-DNA hybrids *in vivo*

For both proposed models of R-loop-dependent dsRNA formation, the displaced single stranded DNA is expected to support antisense transcription by RNA Polymerase to form asRNA. It has been shown that double stranded DNA template with hyper-negative supercoiling **(Parvin & Sharp, 1993)** or with a mismatch in upstream initiation sequence **(Pan & Greenblatt, 1994)** can initiate transcription by RNA Pol II *in vitro* in the presence of only TBP and TFIIB minimal transcription factors. Besides, the formation of RNA-DNA hybrids over double strand break (DSB) ends **(Ohle *et al.*, 2016)** suggests that, similarly, ssDNA of R-loops may serve as a template for transcription and hybrid formation. Yet, the conditions and sequence requirements for transcription over ssDNA *in vivo* are poorly understood. Elegant work was carried out by other colleagues in the lab to better understand this process. They have shown that artificial dsDNA constructs with ssDNA gaps or overhangs, not only support transcription, but also, support RNA-DNA hybrid formation *in vivo* (unpublished data). In general, these results suggest the idea that ssDNA can support transcription. In particular, these findings also support the suggestion that ssDNA can form an RNA-DNA hybrid and so may do a ssDNA in an R-loop, which doesn't exclude the double R-loop formation model. These findings paved the way to further investigate the contribution of R-loops to dsRNA formation.

In the middle of this work, a new publication came to light showing that *in vivo* R-loops act as Pol II promoters that induce antisense transcription (AS) in mammalian cells **(Tan-Wong *et al.*, 2019)**. However, this study didn't try to study the involvement of these AS transcripts in hybrid formation or the role of R-loops in dsRNA generation.

### 4.3.2 Previous ChIP-exo data suggests double R-loop formation over *S. pombe* genome

To test the first model and examine double R-loop formation, a previously in-lab generated R-loop mapping data for RNase H-depleted *rnh1*Δ *rnh201*Δ fission yeast cells (non-crosslinked) using ChIP-exo method **(Ohle *et al.*, 2016),** was analyzed to check the correlation between forward and reverse strand R-loops signals over enriched regions. This was expected

to give an idea about the formation of double R-loops genome-wide. This ChIP-exo data was chosen for analysis as there were only two more mapping data available for fission yeast at the time of analysis. Both of which were generated using classic DRIP-seq **(Bronner *et al.*, 2017; Castel *et al.*, 2014)** and found to lack both resolution and directionality whereas the ChIP-exo revealed to have better quality and higher resolution. Surprisingly, metagenic analysis for this ChIP-exo data, showed bidirectional metagenic signals with closely similar distribution over both forward and reverse DNA strands. Importantly, these signals exhibited overlapping peaks with comparable strengths at the center of the mapping window (Figure 4.5). This analysis suggested a strong correlation between forward and reverse strand signals that when one DNA strand forms an RNA-DNA hybrid, the complementary strand will form another anti-directional hybrid at a similar frequency. This preliminary analysis suggested that double R-loop structures may form widely over fission yeast genome.

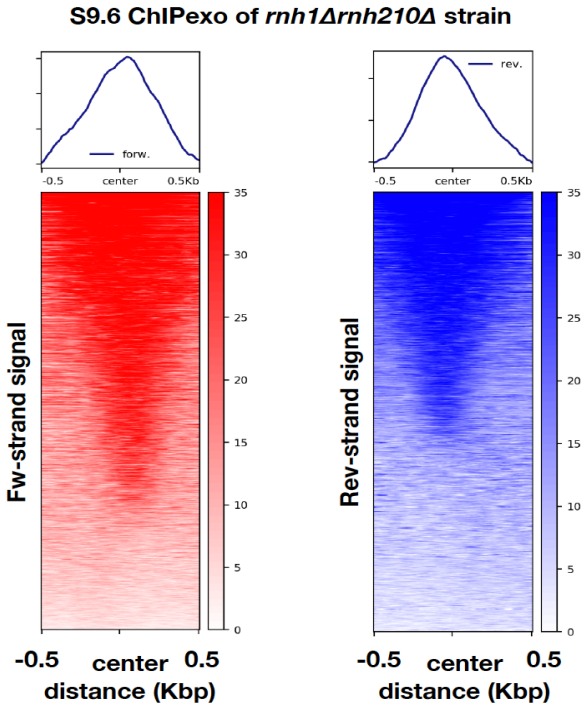



**Figure 4.5. Metagenic analysis for previous S9.6 ChIP-exo data of *rnh1Δ rnh201Δ* yeast cells.** Heat map showing a strong correlation between forward (fw)-strand and reverse (rev)-strand R-loops signals over enriched regions. Analysis reveals metagenic signals of similar distribution, comparable strengths and overlapping central peaks. Data belongs to a previous publication from Fischer lab **(Ohle *et al.*, 2016)**.

Although this was a very promising finding, it was very crucial to confirm the precision and reproducibility of these results using different approaches, especially, knowing that these R-loop mapping experiments weren't controlled by neither endogenous RNase H expression nor exogenous RNase H treatment. In addition, this data belongs to RNase H-depleted (*rnh1*Δ

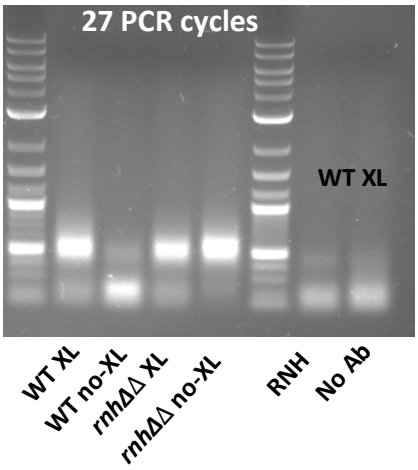
*rnh201*Δ) yeast cells which may have increased R-loop levels and different R-loop formation patterns compared to the normal physiological conditions of the WT cells. Still a major concern about this method and corresponding data is the directionality of signals, which hasn't been confirmed. To ensure reproducibility and consistency of results obtained above, I planned to map R-loops for WT and *rnh1*Δ *rnh201*Δ cells with different methods that maintain high resolution and strand specificity, and, more importantly, enable directional mapping. I also planned to map both of the DNA and the RNA, separately, for formaldehyde crosslinked (XL) and non-crosslinked (no-XL) samples, and to use appropriate controls for validating the data.

### 4.3.3 Quality controls of dsDNA-based ChIP-exo 1 (dsChIP-exo) for R-loop mapping

R-loops profiling can be achieved by either sequencing and mapping the DNA or the RNA strand of the RNA-DNA hybrid captured by S9.6 antibody. I started with mapping the DNA strand as DNA-based R-loop mapping methods have proven to be more robust (produce high yield and show clear signals) compared to RNA-based methods. Another reason for starting with a DNA-based method is that S9.6 Ab has been shown to have considerable affinity to dsRNA which would affect the final results if sequencing is to be done for RNA. DNA-based mapping methods can be either double-stranded or single-stranded DNA based. An example for the former is the original ChIP-exo (ChIP-exo 1) which is based on the ligation of double stranded adaptors to dsDNA. Henceforth, I will call this method as dsChIP-exo.

Regardless of the directionality issue, dsChIP-exo is still a powerful and robust technique. Its data can be used as a standard to which R-loops profiles, generated by current methods, can be compared. It's also important to check reproducibility of the above correlation using the same method. Thus, I found it more convenient to first use this method to map R-loops in the RNase H depleted, *rnh1*Δ *rnh201*Δ*, S. pombe* cells. In addition to *rnh1*Δ *rnh201*Δ yeast cells, I included WT cells for only comparing the final amplification results and quality controls to each other. Moreover, I've conducted the experiment using crosslinked and non-crosslinked samples for each cell type. As controls, and to confirm the specificity of the results, I included RNase H-treated and no-antibody control samples. Importantly, I adopted chromatin immunoprecipitation conditions, which means I used whole cell lysate with chromatin preparation instead of nucleic acids extract as input for immunoprecipitation using S9.6 antibody (see discussion). For fragmentation, I used sonication instead of enzymatic digestion to enhance the resolution and avoid the enrichment of long fragments. This was guided by

recommendations of a previous study to use chromatin preparation in conjunction with sonication **(Halasz *et al.*, 2017)**.




| Sample | Q5 PCR cycles |
|---|---|
| WT XL | 15 |
| WT no-XL | 20 |
| *rnhΔΔ* XL | 15 |
| *rnhΔΔ* no-XL | 15 |
| WT XL + RNH | 20 |
| WT XL, no-Ab | 20 |

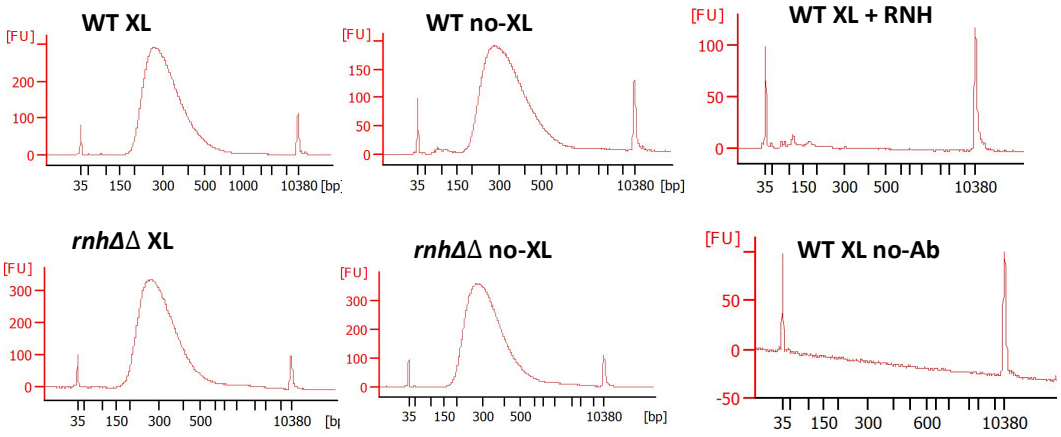



**Figure 4.6. Quality controls for S9.6 libraries of dsChIP-exo.** Agarose gel analysis after check PCR amplification (top) and bioanalyzer profiles of final libraries (bottom). Check PCR for libraries of crosslinked (XL) and non-crosslinked (no-XL) WT and *rnh1Δ rnh201Δ* (*rnhΔΔ*) cells shows robust amplification for all the libraries except the non-crosslinked WT, and RNase H (RNH)-treated and no-Ab control samples. Bioanalyzer profiles shows empty RNH-treated and no-Ab controls. The table shows number of PCR cycles for final amplification before bioanalyzer analysis.

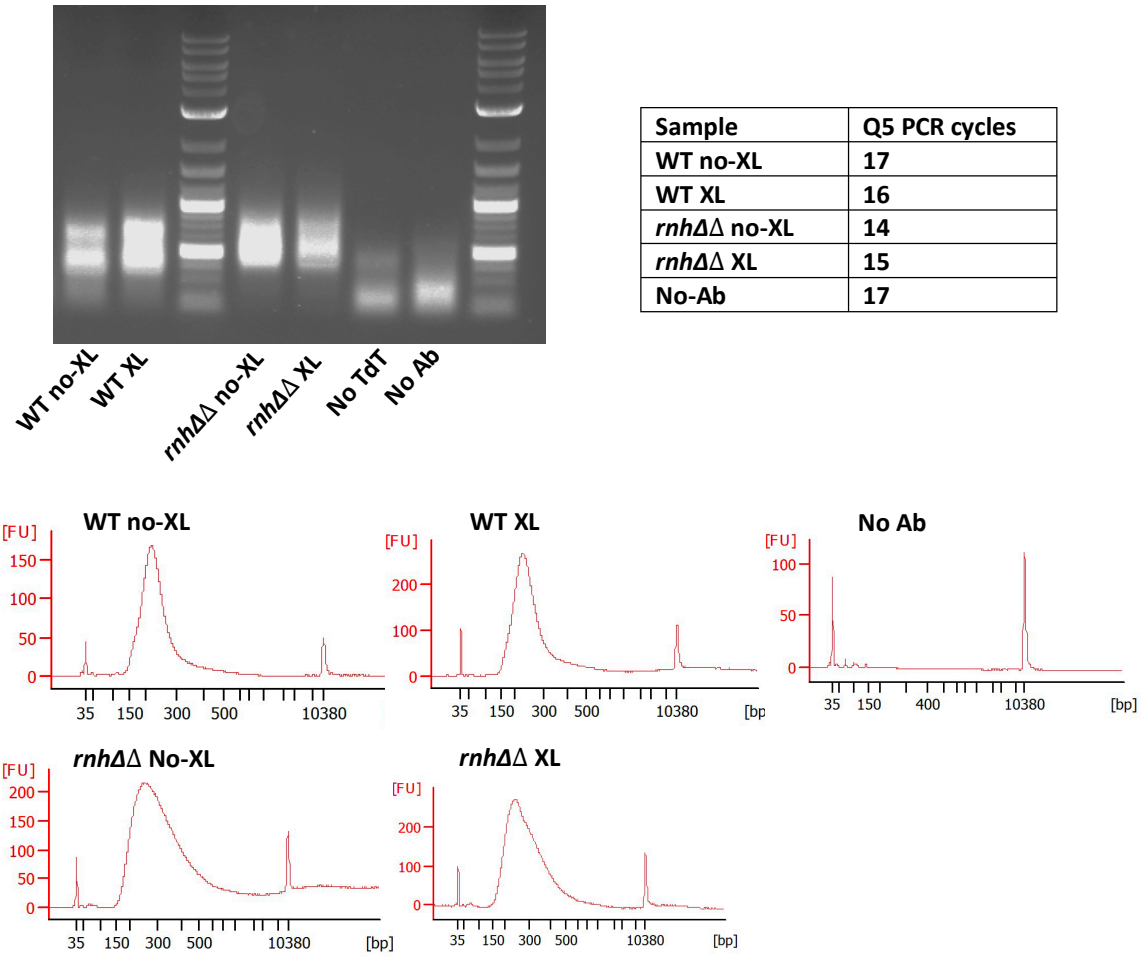
Surprisingly, both of the RNase H-treated and no-antibody controls were almost empty of PCR amplification products. This was very clear after agarose gel separation and bioanalyzer analysis as shown in Figure 4.6. For consistency, the control libraries were amplified using the highest PCR cycle number used for experimental samples. This provides a quiet strong evidence that the amplification products are specific for R-loop forming sequences. Apparently, the library of non-crosslinked WT cells didn't produce enough amplification products after the first *Taq* PCR and agarose gel separation (Figure 4.6). For this reason, I used 5 extra PCR cycles for Q5 amplification of this sample to produce a bioanalyzer signal

comparable to the other libraries. Different from WT cell sample, the PCR amplification for the crosslinked *rnh1*Δ *rnh201*Δ cell sample was weaker than that for the non-crosslinked sample of the same cell type. This suggests that crosslinking may induce some variations in the mappings. However, it's not clear why crosslinking shows different results for WT and *rnh1*Δ *rnh201*Δ cells (next section and discussion).

### 4.3.4 Quality controls of ssDNA-based ChIP-exo-TT (ssChIP-exo) as a new R-loop mapping method

It is quite clear how original ChIP-exo (ChIP-exo 1) reactions work with protein-bound DNA (Chapter 3). However, it's not clear how these reactions, especially double stranded adaptor ligation, work for R-loops or RNA-DNA hybrids. The ligation step requires dsDNA ends, a condition which may not be available for all captured R-loops (see discussion), which raises some concerns regarding directionality. The prerequisite for dsDNA ends may also impose limitations on sensitivity. This situation applies to all dsDNA-based DRIP methods without exception. This is an enough reason why ssDRIP-seq, a single-stranded DNA library preparation and ligation-based method, has been introduced and used to map R-loops in *Arabidopsis* **(Xu *et al.*, 2017)**. For the first ligation step of this method, a single stranded adaptor is ligated to the 3' end of S9.6-immunoprecipitated DNA. Surprisingly, the authors reported that this method captured only the DNA strand of the RNA-DNA hybrid but not the free displaced ssDNA, which confirms the directionality of the ssDRIP-seq. Despite its directionality, the ssDRIP-seq may suffer from low resolution. To maintain high resolution, sensitivity and directionality at the same time, I sought to harness the ChIP-exo, but, combined with a sensitive ssDNA-based capturing method. To approach this, I employed the ChIP-exo-TT, a modified version of ChIP-exo that I have developed during this project and that uses the principle of poly-deoxynucleotide tailing for DNA 3' ends (Chapter 3). For this method, the lambda exonuclease treatment is done directly after the polishing & kinase reaction while the 1$^{st}$ adaptor ligation step is omitted, but, instead, a poly-dA tailing step is done for the 3' end of DNA. Poly-dA-tailed DNA is then extended using an oligo-dT primer for generation of dsDNA. I found ChIP-exo-TT to have higher sensitivity compared to the dsChIP-exo due to the increased efficiency of tailing compared to ligation (Chapter 3). Hereafter, I will call ChIP-exo-TT as ssChIP-exo for simple differentiation from dsChIP-exo.

To compare R-loops profiles and results of ssChIP-exo to those of dsChIP-exo, I used the same crosslinked (XL) and non-crosslinked (no-XL) *rnh1*Δ *rnh201*Δ yeast cells. This was required to test the reproducibility of the correlation between the forward-strand and reverse-strand R-loops signals. Importantly, this time, I used WT cells for R-loop mapping in order to study the same correlation and to examine the double R-loop model using physiological conditions. Comparing WT R-loop profiles against those of RNase H-depleted cells would be helpful to distinguish real R-loop signals by sensitivity to WT RNase H levels. This would present a better natural control compared to RNase H-treatment which I already used for the previous dsChIP-exo, and indeed confirmed the specificity of the total PCR signal (see discussion).




| Sample | Q5 PCR cycles |
|---|---|
| WT no-XL | 17 |
| WT XL | 16 |
| *rnhΔΔ* no-XL | 14 |
| *rnhΔΔ* XL | 15 |
| No-Ab | 17 |

**Figure 4.7. Quality controls for S9.6 libraries of ssChIP-exo.** Agarose gel analysis after check PCR (top) and bioanalyzer profiles for final libraries (bottom), showing robust amplification for all the samples except no-Ab and no Terminal deoxy-Transferase (no-TdT) controls. The table shows number of PCR cycles for final amplification before bioanalyzer analysis.

Similar to results obtained by dsChIP-exo, PCR signal strengths were different between crosslinked and non-crosslinked samples of the same strain, with the weakest amplification level for non-crosslinked WT (Figure 4.7). While crosslinking stregnthened the cumulative signal of WT, it weakend the signal of *rnh1*Δ *rnh201*Δ cells. Importantly, the fact that both dsChIP-exo and ssChIP-exo methods showed the same results confirms the consistency of methods and reproducibility of results. In general, and regardless of the final effect, these observations match know facts that crosslinking induces some variations in R-loop mapping signals **(Halasz *et al.*, 2017)**. The difference in results obtained for WT and *rnh1*Δ *rnh201*Δ cells can be explained in lights of previous studies which reported different effects for formaldehyde fixation **(Halasz *et al.*, 2017)**. Formaldehyde fixation was reported to induce conformational changes and lead to denaturation of dsDNA **(McGhee & Von Hippel, 1977)**, which can induce R-loop formation in presence of RNA. In addition, formaldehyde fixation may also crosslink different proteins to DNA or R-loop sequences making some R-loops inaccessible for the antibody. It has been reported that fixation partly affects recognition of R-loops by S9.6 **(Legros *et al.*, 2014)**. Comparing the effect of crosslinking on WT (which is profound) to those of *rnh1*Δ *rnh201*Δ cells (which is weak), it seems that the R-loop induction effect overcomes the antibody-recognition blocking effect. In fact, it can't be excluded that crosslinking may exert both effects on R-loop signals of the same sample, but the final net outcome isn't clear (see next sections and discussion).

Strikingly, the no-Terminal deoxy-Transferase (no-TdT) control library was hardly amplified, which confirms the dependence of the signal on poly-dA tail added by the tailing enzyme, but not on genomic dA tracts that may be present in the DNA template. This final issue doesn't matter that far as long as the mapping is specific which is confirmed by the PCR signal-free no-antibody control (Figure 4.7). The absence of PCR signal in the no-Ab control, confirms the specificity of the PCR signal to amplified R-loop forming sequences. In this experiment, I didn't use RNase H-treatment control as I compensated this by the use of WT against RNase H-depleted cells for comparing signals after mapping. In addition, comparing the ssChIP-exo mapping signals to those of the dsChIP-exo for which general specificity was confirmed by the RNase H-treatment, should reveal the specificity of the ssChIP-exo mapping (Figures 4.9 & 4.11).

### 4.3.5 Quality controls of SMART-DRIPc as a new RNA-based R-loop mapping method

ssChIP-exo is supposed to be a directional method as it is based on single-stranded DNA capturing approach that should capture only the DNA strand of the hybrid. However, it is hard to confirm at this stage that the displaced ssDNA will be completely removed during sonication and washing steps as previously suggested **(Crossley *et al.*, 2020; Xu *et al.*, 2017)**. If recovered, this ssDNA will get captured and sequenced to produce a false anti-directional signal and compromise the directionality. To completely exclude this possibility, I decided to use an RNA-based R-loop mapping method to map only the RNA strand adopting ChIP- and S9.6-dependent DRIP-like workflow similar to dsChIP-exo and ssChIP-exo.

The RNA-based DRIPc was the best candidate to use. However, a major caveat for this method is that S9.6 has considerable affinity to dsRNA. This has been shown to affect R-loop mapping, especially in fission yeast **(Hartono *et al.*, 2018)**. In addition, the signal resulting from mapping RNA may have low resolution due to the free ssRNA. Therefore, the sample has to be treated with an enzyme that would digest the dsRNA while another enzyme should be used for digesting the free ssRNA to enhance the resolution. Due to the expected low RNA yield from DRIPc experiment, a high sensitivity library preparation method is required to compensate for the loss, and show robust enrichment comparable to DNA mapping methods.

First of all, I carried out some pilot DRIPc experiments to the point of elution and input amount estimation in order to determine the amount of RNA obtained from a single experiment and, also, to decide the optimal conditions for the RNase H treatment. Accordingly, I tried two ways for the RNase H treatment, either after S9.6 IP and during on-beads incubation, or after final RNA extraction (in-solution). Unfortunately, for a chromatin-based DRIPc, this reaction can't be done directly after the lysis (before adding the antibody) as the chemical detergents used for lysis may inactivate the RNase H. Surprisingly, I found the RNase H treatment to be more efficient when done while the chromatin/RNA is still bead-bound than when done after phenol chloroform extraction and ethanol precipitation. This was revealed by direct RNA concentration measurements using the Qubit (Table 4.1). A possible explanation for this observation is that RNA extraction steps may affect the integrity of some R-loops and produce ssRNA that is resistant to RNase H. Recently, it has been found that ethanol precipitation can lead to the resolution of some artificial R-loops **(Crossley *et al.*, 2020)**. Regardless of the

reason behind this, some RNA was still detected using Qubit even after on-bead RNase H treatment (Table 4.1) (see discussion).

**Table 4.1. RNase H treatment is more efficient during on-bead incubation**. RNA concentration measurements after an RNA IP using S9.6. No-antibody is out of range, some RNA is still detected after RNase H (RNH) treatment. Measurement was done using qubit high sensitivity kit. Final volume for each sample was 25 µl.

| Sample | Conc. (ng/µL) |
|---|---|
| No-antibody | Out of range |
| RNH on beads | 2.4 |
| RNH after extraction | 5.12 |
| IP sample | 7.02 |

| Sample | Conc. (ng/µL) |
|---|---|
| RNH on beads | 1.3 |
| IP sample | 3.86 |

The next step was to test suitable enzymes that can digest dsRNA and free ssRNA but not RNA of R-loops. It has been reported that RNase A can be used in high salt concentration reaction conditions to digest dsRNA without affecting RNA strand in the RNA-DNA hybrid **(Z. Z. Zhang *et al.*, 2015)**. In my hands, using RNase A in high salt led to complete loss of RNA (RNA undetected). It has been reported that the use of RNase A which high salt cuases substantial loss and under-representation of R-loop signals **(Halasz *et al.*, 2017; F. Konig *et al.*, 2017; Sanz & Chedin, 2019)** depending on factors such as buffer conditions, incubation time and salt concentration **(Halasz *et al.*, 2017)**. This loss produceed significant changes in R-loop mapping results **(Halasz *et al.*, 2017)**. Considerably, the use of high salt, especially with on-bead incubation may affect the stability of the bead-antibody immunocomplexes, deplete IPed material and lead to loss of signals. Also, RNase A was reported to have strong DNA-binding **(Benore-Parsons & Ayoub, 1997; Dona & Houseley, 2014; Halasz *et al.*, 2017)** and detectable RNA-DNA hybrid-degrading activity **(Ausubel *et al.*, 2011; Smolka *et al.*, 2021)** which can definitely contribute to loss of signals. For these reasons, RNase A is unreliable and suboptimal control for R-loop mapping experiments **(Smolka *et al.*, 2021)**.

As an alternative for RNase A, I used RNase III (RNIII) which splices dsRNA into small fragments. This strategy has been previously used for fission yeast DRIPc by **Hartono *et al.*, 2018**. Moreover, I incorporated a second treatment step using P1 nuclease to digest ssRNA and ssDNA. P1 nuclease is a widely used enzyme in hybridization and nuclease protection assays to digest single-stranded nucleic acids. Importantly, treating with P1 would be helpful to confirm that ssRNA is digested and exclude the capture and sequencing for free mRNA. After

successfully detecting S9.6 IP RNA, even with three sequential enzymatic treatments, the next hurdle was to optimize a library preparation method for low-input RNA. The SMART method was the best candidate for this mission. Similar to DNA-based SMART used for the ChIP-exo libraries preparation, the RNA was planned to be poly-A tailed using a poly-A Polymerase (PAP), then reverse transcribed for first strand cDNA synthesis and template switching. Before starting the A-tailing, I considered the fact that the different nucleases used after the IP (RNase III or P1), may affect the phosphorylation state of the 3' and 5' ends of RNA, which may affect either the A-tailing or the template switching or both. I checked this possibility be treating the RNA with different modification enzymes to figure the best conditions (Figure S2 and Table S5). Amazingly, PCR amplification was robust for all conditions used except the no-tailing control which showed very weak amplification though. This proves the signal is dependent on the artificially added poly-A tail. Slight amplification of the no-tailing control may be expected knowing that the immunoprecipitated RNA may have poly A-tracts (Figure S2). This wouldn't affect the results as long as the ChIP is specific enough, which will be shown to be true in the next parts.

| Sample | Q5 PCR cycles |
|---|---|
| IP | 17 |
| IP + RNIII + P1 | 15 |
| No-Ab | 17 |
| IP + RNH | 20 |

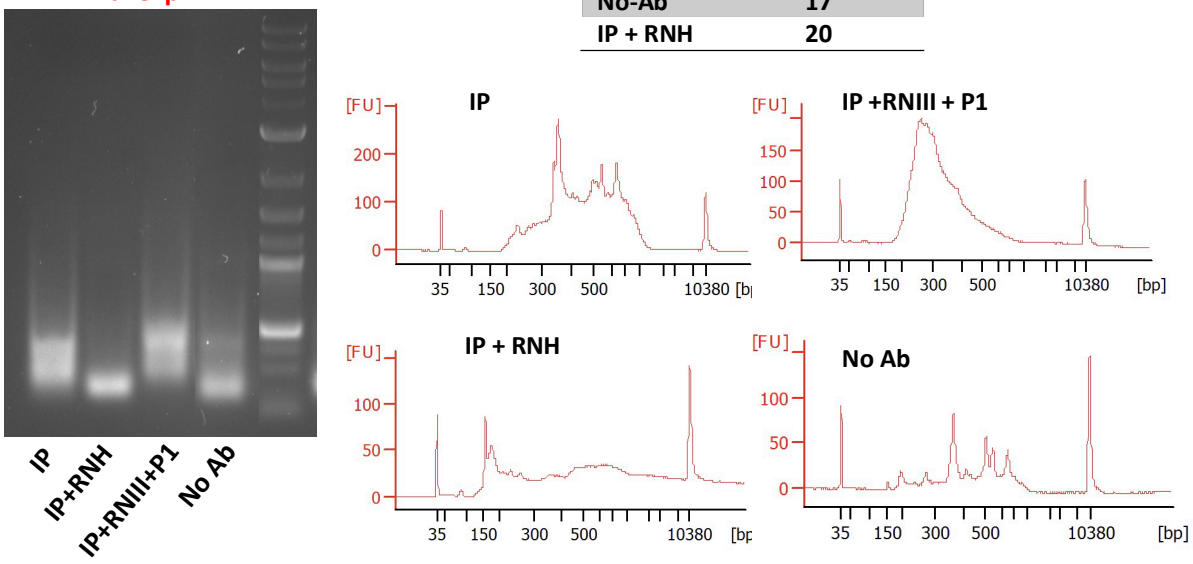



**Figure 4.8. Quality controls for S9.6 libraries of SMART-DRIPc.** Agarose gel separation after check PCR amplification (left) and bioanalyzer profiles for the final libraries (right) of non-crosslinked *rnh1Δ rnh201Δ* cells, showing robust amplification for untreated IP, RNase III (RNIII) & nuclease P1 (P1)-treated, but not RNase H (RNH)-treated IP or no-Ab control samples. The table shows number of PCR cycles for final amplification before bioanalyzer analysis. Note that only 22 PCR cycles were used for check PCR.

For the final SMART-DRIPc mapping experiment, I included four different samples with controls; the untreated immunoprecipitate (IP), the RNase III plus P1 nuclease (RNIII/P1)-treated IP, the RNase H-treated IP and no-Ab control. After elution and RNA extraction, I treated the RNA with Shrimp Alkaline Phosphatase (SAP) to dephosphorylate the 3' end of RNA which may result after treating with P1 (leaves 3'-P RNA ends) which may block the tailing reaction. For the SMART step, I processed the RNA following single-tube workflow (no purifications). Both of the *Taq* PCR and Q5 final amplifications showed weak amplification for the RNase H-treated and no-Ab control compared to IP samples (Figure 4.8). This means in general the PCR signal is specific for R-loop-forming sequences. The amplification in the RNase H control may be due to enhanced sensitivity of SMART or due to resistance of some R-loops to RNase H **(Chédin *et al.*, 2021; Crossley *et al.*, 2020; K. Wang *et al.*, 2021)**.

### 4.3.6 Reproducibility of the current S9.6 profiles and overlap among signals

To have a look over the quality and reliability of the data produced by my mapping methods (hereafter called current methods), I planned to compare my data against publicly available data. For this, I retrieved public R-loop mapping data for fission yeast through NCBI GEO to be analyzed the same way with my data. I classified these data based on the sequenced template into DNA-based and RNA-based data. The DNA-based data included previous ChIP-exo data from Fischer lab **(Ohle *et al.*, 2016)** (hereafter called external ChIP-exo) and two different DRIP data from two different labs, Martienssen lab **(Castel *et al.*, 2014)** (hereafter called external DRIP) and Halic lab **(Bronner *et al.*, 2017)** (hereafter called DRIP). Halic lab DRIP data was used only for comparing signal localization using genome browser, but excluded from further analysis.

The RNA-based data included only one data generated by DRIPc (hereafter called external DRIPc) through a collaboration between Chédin and Vanoosthuyse labs **(Hartono *et al.*, 2018)**. After analyzing all the data in parallel using the same bioinformatic procedures, I compared R-loops profiles generated by similar methods together (either DNA- or RNA-based). For instance, I compared SMART-DRIPc to external DRIPc data and, separately, the rest of DNA-based data together. Beside considering the spatial distribution (overlapping) of signals among the different R-loops profiles as an important parameter for the comparison, I considered four more factors which are resolution, directionality, sensitivity, and enrichment of signals over genomic regions already identified as hot spots for R-loops in fission yeast.

#### *4.3.6.1 Overlap of DNA-based methods signals*

After trimming adaptors and filtering short reads, reads were aligned to fission yeast genome (ASM294v2, PomBase), and positive signal peaks were predicted for the forward and reverse strand, separately, using MACS2. Examination for the numbers of signal peaks produced by different methods, revealed that dsChIP-exo produced 3220 and 3205 signal peaks for non-crosslinked and crosslinked *rnh1*Δ *rnh201*Δ samples, respectively, representing the highest number of signal peaks detected (Figure 4.9A). Comparatively, ssChIP-exo produced lower number of signal peaks for the *rnh1*Δ *rnh201*Δ samples, with 1994 and 1688 signal peaks for the non-crosslinked and crosslinked cells, respectively, representing nearly half the number of signal peaks detected by dsChIP-exo (Figure 4.9A).

It may be tricky to draw a conclusion based on the difference in number of signal peaks captured by the dsChIP-exo and ssChIP-exo. This number may denote either a difference in signals locations (number of enriched regions) or a difference in directionality, since the signal peaks are predicted for the forward and reverse strands, separately. Direct inspection for the signals using the genome browser showed a very strong agreement (near optimal spatial overlap) between dsChIP-exo and ssChIP-exo signals (Figure 4.9B). Importantly, I noticed a difference in signals directionality, which may explain the reason behind the difference in number of signal peaks detected by both methods (Figure 4.9B). Amazingly, 90% of ssChIP-exo signal peaks overlapped with 55% of dsChIP-exo signal peaks for no-XL *rnh1*Δ *rnh201*Δ samples (Figure 4.9A). Close overlap ratios were revealed for XL *rnh1*Δ *rnh201*Δ samples of ssChIP-exo and dsChIP-exo which detected 1508 signal peaks in common, representing an overlap rate of ~89% and ~47%, respectively, confirming the consistency and reproducibility of mapping profiles. It's important to note that maximum overlap from dsChIP-exo should be around 50% of signal peaks. The reason is that this method may not be directional, which means each locus is representd by two signals (i.e., two signals from dsChIP-exo is equivalent to one signal form ssChIP-exo for a specific locus) (see next sections and discussion).

Comparing the overlap of signals captured by dsChIP-exo (which detected the highest number of signal peaks) to those of the external DNA-based mapping methods, I found that 1412 out of 3220 dsChIP-exo signal peaks (44%) overlapped with signal peaks of both external DRIP and external ChIP-exo. Although the overlap between external DRIP and external ChIP-exo data was only 343 signal peaks, dsChIP-exo showed an overlap of 872 and 794 signal

peaks with each data, respectively, and separately (Figure 4.9A). While the overlapping signal peaks confirm the consistency of results, the big fraction of non-overlapping signal peaks in dsChIP-exo may be explained by increased sensitivity (robustness) of this method, technical differences among methods or genotype variation among yeast cells. Regarding the latest point, the external DRIP was conducted for WT cells while the external ChIP-exo was conducted for *rnh1Δ rnh2Δ* cells engineered to express an exogenous restriction enzyme.

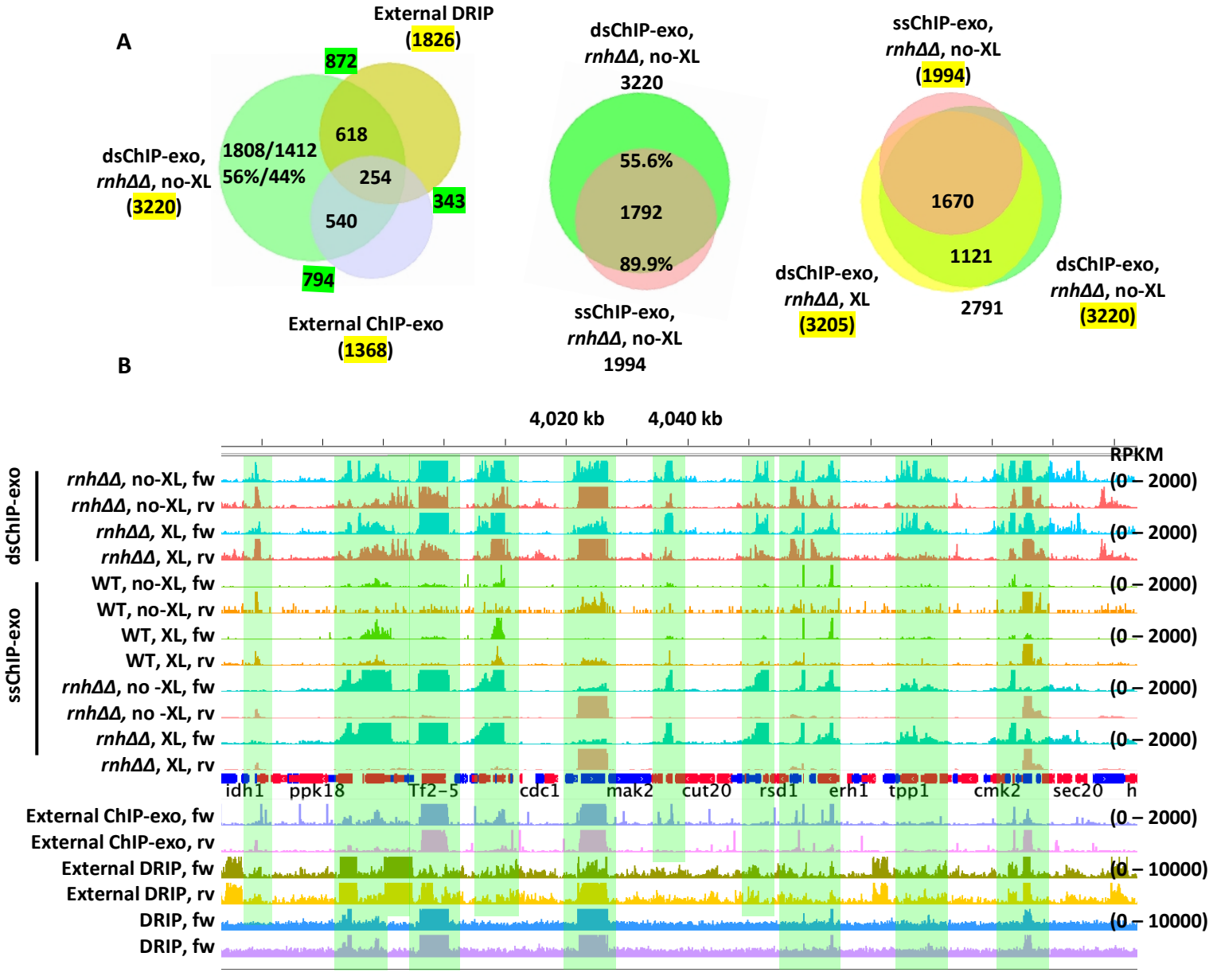



**Figure 4.9. Overlap of dsChIP-exo and ssChIP-exo signals with those of other DNA-based methods. (A)** Venn diagram analysis showing number of and overlap of signal peaks captured by current and external methods. **(B)** snap shot of genome browser showing S9.6 profiles of different methods over a genomic region of *S. pombe*. Highlighted regions represent overlapping signals. Scaling was adjusted to reveal weak- and medium-strength signals and, at the same time, to distinguish real signals from background.

Finally, the consistent decrease in number of signal peaks detected with my methods for crosslinked compared to non-crosslinked *rnh1*Δ *rnh201*Δ cells (Figures 4.9 & 4.11) is consistent with the strength of PCR signals of their libraries (Figures 4.6 & 4.7). This observation is also consistent with previous findings that formaldehyde crosslinking weakens the recognition of R-loops by S9.6 antibody *in vitro* **(Legros *et al.*, 2014)**. However, this pattern was reversed for WT cells where a PCR signal wasn't detected for non-crosslinked WT cells of dsChIP-exo (Figures 4.6 & 4.7). This suggests that crosslinking may affect the enrichment of signals based on some unknown factors (Section 4.3.4 and discussion).

#### *4.3.6.2 Overlap of RNA-based methods signals*

Surprisingly, the SMART-DRIPc produced 2952 and 1598 signal peaks for the untreated and RNIII/P1-treated *rnh1*Δ *rnh201*Δ samples, respectively, showing high sensitivity similar to the dsChIP-exo (Figure 4.10A). It is quite reasonable that processing the RNA with two extra enzymatic treatments (RNase III and P1 nuclease) would affect the signal peak numbers and strengths. However, the reason behind this effect isn't clear as it may be due to increased directionality or, otherwise, due to decreased efficiency of the mapping, a possible result for using multiple nucleases that might have depleted the signals. Inspecting the signals over the genome browser showed a very strong overlap among the signals of the untreated and RNIII/P1-treated SMART-DRIPc samples, but weaker signals for the latter (Figure 4.10B). There was no difference in directionality suggesting that the use of these enzymes may have decreased the efficiency of the mapping.

Comparing the overlap of signal peaks of the RNA-based methods showed that around 75% of the RNIII/P1-treated SMART-DRIPc samples signal peaks overlapped with ~40% of those of untreated SMART-DRIPc samples (Figure 4.10A). Each of the untreated and RNIII/P1-treated SMART-DRIPc samples showed a similar overlap of ~500 signal peaks with external DRIPc data for *rnh1*Δ *rnh201*Δ samples. Importantly, more than 75% (1198 of 1589) of the RNIII/P1-treated SMART-DRIPc samples signal peaks overlapped with both of the untreated SMART-DRIPc and the external DRIPc samples signal peaks which confirms the reproducibility of the mapping of the RNIII/P1-treated *rnh1*Δ *rnh201*Δ cells (Figure 4.10A). Inspecting R-loops profiles over genome browser showed that every single signal captured by SMART-DRIPc coincide with a signal from external DRIPc. The same applies for untreated and RNIII/P1treated SMART-DRIPc samples (Figure 4.10B). The main reason for the non-

overlapping fraction of signal peaks may be the overrepresentation and underrepresentation of some signals among methods.

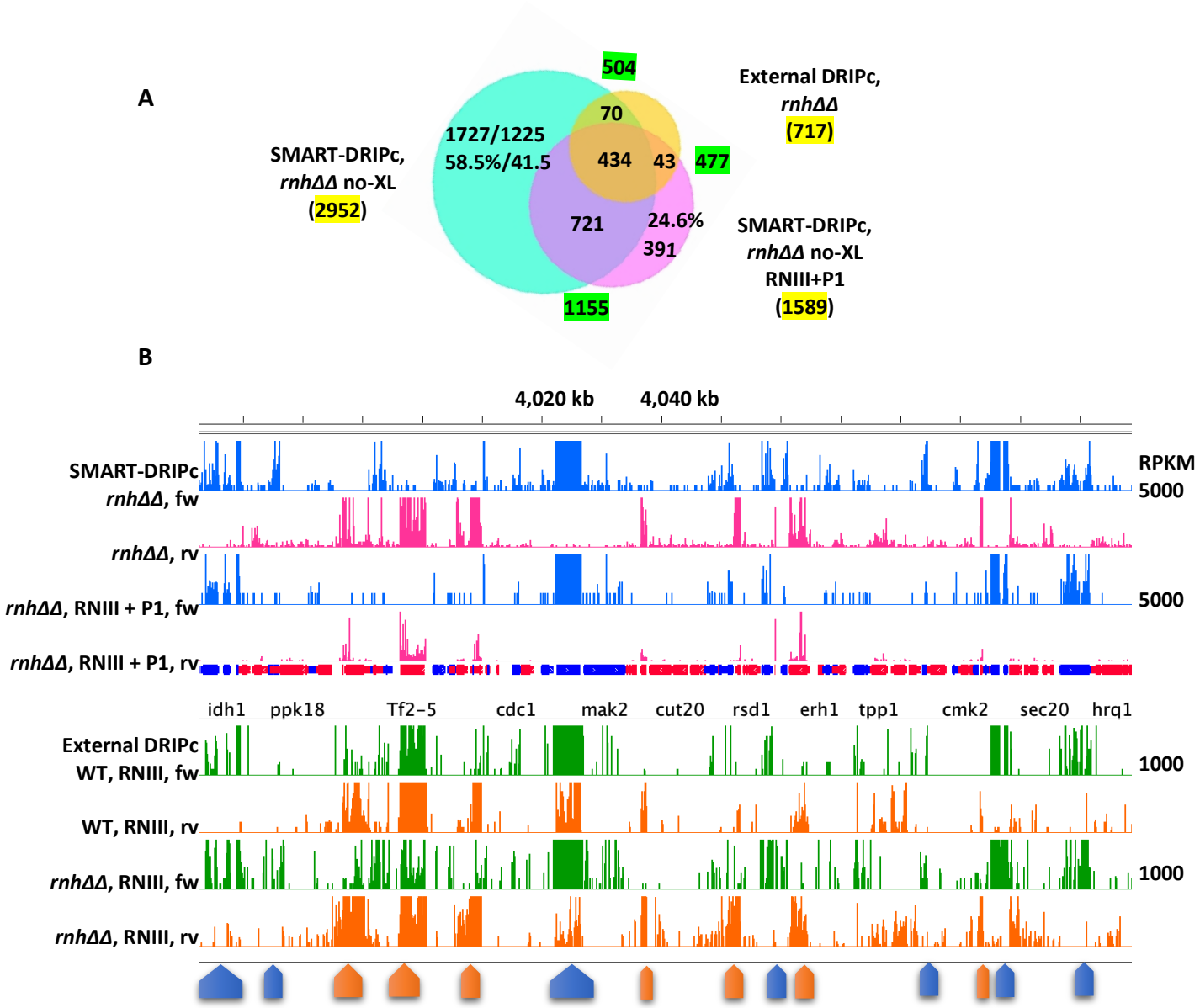



**Figure 4.10. Overlap of SMART-DRIPc signals with those of external DRIPc. (A)** Venn diagram analysis showing number and overlap of signal peaks. **(B)** snap shot of genome browser showing S9.6 profiles of different methods over a genomic region of *S. pombe*. Blue and red shapes point to overlapping forward-strand and reverse-strand signals, respectively. Scaling was adjusted to differentiate positive signals form background and to reveal all signals.

#### *4.3.6.3 Overlap of DNA and RNA-based methods signals*

The greatest degree of overlap detected was seen for the dsChIP-exo and ssChIP-exo signal peaks (Figures 4.9 & 4.11). However, this overlap was lower between each of my ChIP-exo methods, separately, and the SMART-DRIPc method signal peaks. Notably, only 1439 out of 2952 (50%) of SMART-DRIPc signal peaks overlapped with ~45% signal peaks of dsChIP-exo (Figure 4.11). This degree of overlap is relatively very strong compared to previous reports about overlap of signal peaks detected by different R-loop mapping methods. For example, an overlap of only 27.5% and 38.5% was reported between DRIP and R-ChIP signal peaks for K562 human cells **(L. Chen *et al.*, 2017)** while Proudfoot lab reported an overlap of only 25% of DRIPc signal peaks for Ntera2 cells with RDIP signal peaks of Hela cells **(Tan-Wong *et al.*, 2019)**. It's important to note that for the second study, another reason for this partial overlap, beside the technical differences, may be the use of different cells that may have different expression profiles. All together, these observations confirm that the variability in number of signal peaks detected by different methods is an inevitable outcome. This also confirms the fact that the more the technical differences between methods, the lower the overlap between their signal peaks.

Amazingly, more than 90% of the ssChIP-exo signal peaks overlapped with signal peaks form the SMART-DRIPc and dsChIP-exo methods, referring to the reliability of mapping results obtained by the ssChIP-exo as it captures features from different methods. Strikingly, the maximum overlap was observed between signal peaks of crosslinked and non-crosslinked samples for the same method, confirming the reproducibility of results of each method (Figure 4.11). Again, this suggests that similar methods produce closer results while adopting different techniques decreases the overlap.

Finally, the number of signal peaks detected by either the dsChIP-exo or ssChIP-exo for crosslinked is lower than those for the non-crosslinked *rnh1*Δ *rnh201*Δ cells. This confirms the reproducibility of results obtained by both the dsChIP-exo and ssChIP-exo under the different conditions used. Moreover, the overlap of current DNA and RNA-based methods signal peaks was more pronounced compared to the overlap between those of current and external methods. An overlap of ~87% was detected between XL and no-XL *rnh1*Δ *rnh201*Δ cells of dsChIP-exo, while an overlap of 87% and 73%, respectively, was detected for XL and no-XL *rnh1*Δ

*rnh201*Δ cells of ssChIP-exo (Figure 4.11). This strongly evidences for the reproducibility and consistency of data obtained by my methods.

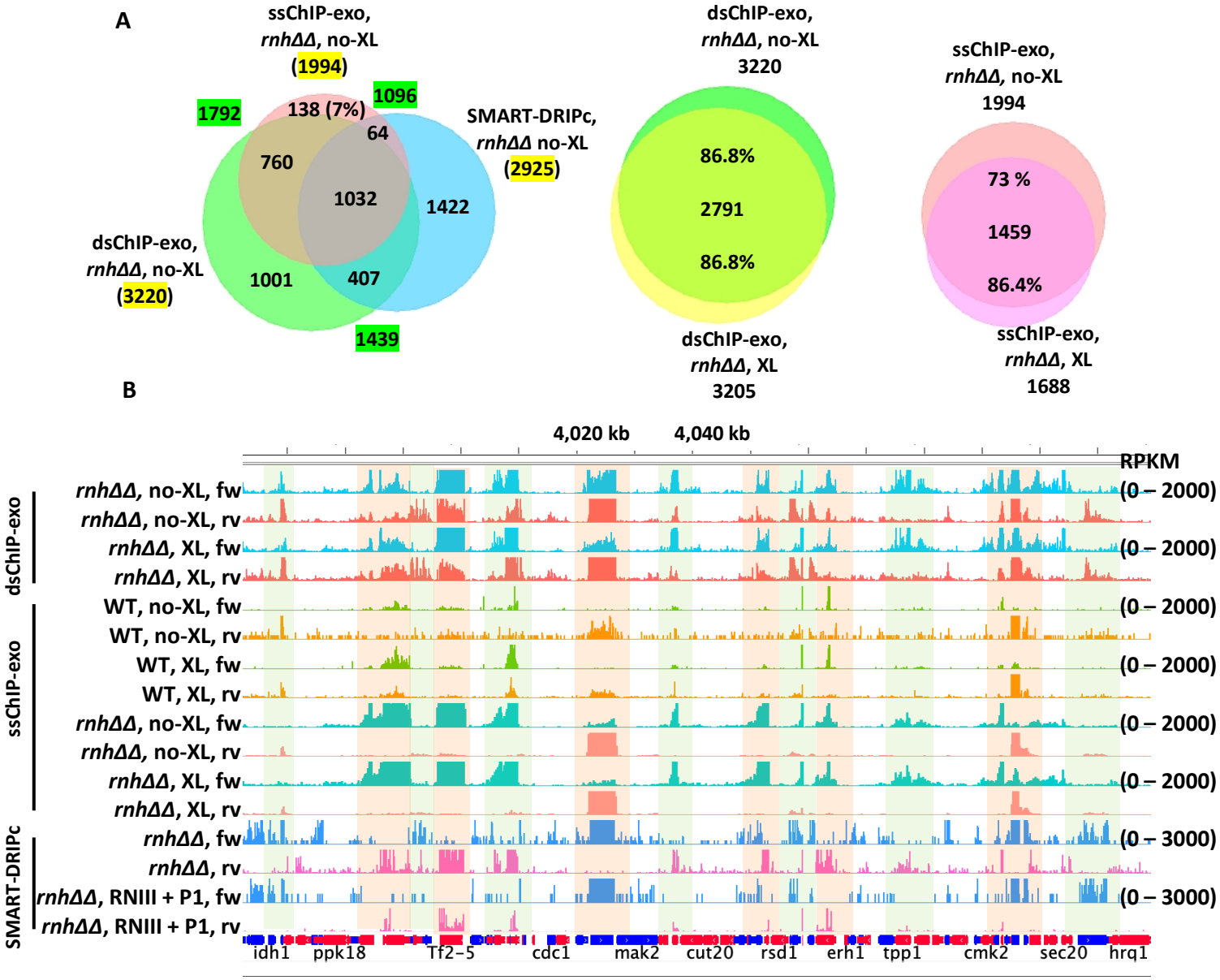



**Figure 4.11. Overlap of signals of all the current DNA- and RNA-based mapping method. (A)** Venn diagrams showing number and overlap of signal peaks for current methods. **(B)** snap shot of genome browser showing S9.6 profiles of different methods over a genomic region of *S. pombe*.

### 4.3.7 Higher resolution and lower noise in current S9.6 profiles

To further confirm the quality of my data and results, I compared the resolution of signals produced by different methods over tDNA clusters at the boundaries of pericentromeric heterochromatin which is a gene-dense region that can clearly show difference in resolution. Surprisingly, the signals detected by the current methods over these loci consistently demarcate the genes bodies, leaving empty gaps over the intergenic regions. On the other hand, the signals of external methods, with different extents, were either absent or extending over the heterochromatin without a distinction for the different genes. Remarkably, signals of the dsChIP-exo and ssChIP-exo showed the highest resolution and the lowest noise to signal ratio (lowest background) compared to the external DRIP and external ChIP-exo data (Figure 4.12). This can be also clearly seen in Figures 4.9 & 4.10 showing R-loop signals over another genomic region. This high resolution and low noise, shown in all the figures in this chapter, is what I was exactly aiming at, for specific investigations in the next chapter.

Remarkably, treating with P1 nuclease didn't seem to have an effect on resolution as it affected signals strengths but not signals widths, implying that the RNA-based R-loop mapping using my experimental conditions has inherent high resolution and doesn't require specific enzymatic treatments (Figure 4.12). Again, it should be noted that sequential treatment for RNA with different enzymes may have affected the signals enrichment. Notably, there was no difference in signal resolution for the crosslinked and non-crosslinked samples for the same method and the same genotype.

Technically, it's possible that the untreated and RNIII/P1-treated RNA may have different ends that can exhibit some bias for the subsequent SMART library preparation steps. If true, this may show differences in the number and strength of signals knowing that all the samples were processed in parallel from one common liquid culture till the SMART step. However, these speculations need to be tested. Collectively, my data shows better quality and enhanced resolution with minimal background compared to other data. It could be argued that this high resolution and minimal background could have come at the expense of lost real signals (as a result of underrepresentation of real signals). However, the strong overlap with external data and that fact that each of my methods captured a much higher number of peaks compared to previous methods doesn't support this assumption. The above overlap analyses suggest that each method may have a bias to capture a specific category of R-loops that may represent a

unique cluster for each method. This explains the absence of some signals in my data (or a specific data) compared to others. Further analysis may be required to understand the features of these R-loop clusters.

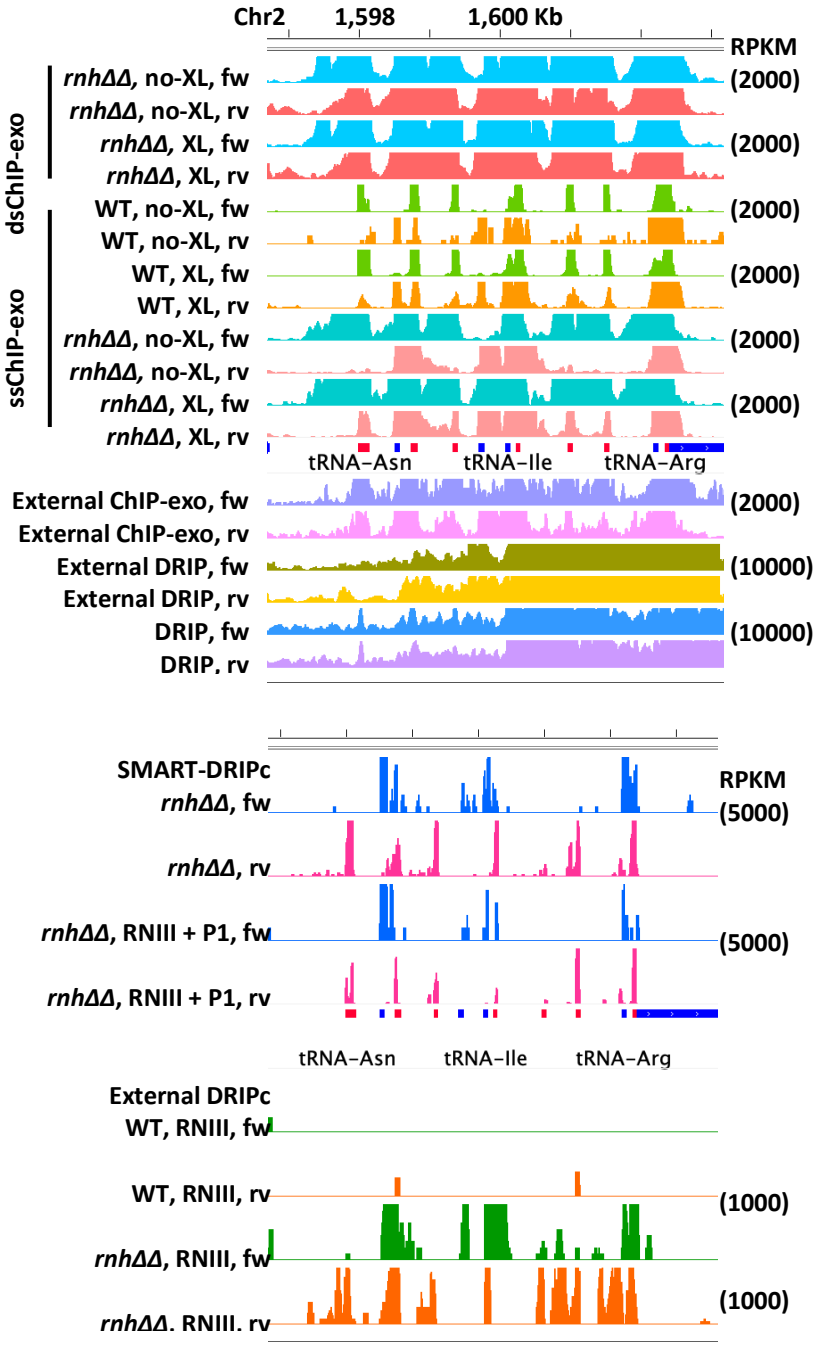



**Figure 4.12. Enhanced resolution of signals detected by current methods.** Snap shot of genome browser showing S9.6 profiles of different methods over pericentromeric tDNA. Compared to external methods, signals of current methods have high resolution as they distinguish tRNA gene bodies from intergenic regions. No background noise is detected for my methods even with low scaling.

### 4.3.8 Signals detected by current methods map to R-loop hot spots of fission yeast

To check the reliability of my data and the specificity of mapped signals, I examined the genomic distribution of signals to test whether enriched loci coincide with known R-loop formation hot spots in fission yeast. Strikingly, strong signals were detected over different non-coding genes such as RNA Pol III-transcribed 5S-rRNA, all tRNAs, and small nuclear (snu) and small nucleolar (sno) RNA genes (Figure 4.13). The strongest signals ever, were found to form over RNA Pol I-transcribed ribosomal RNA genes without exception. These include mitochondrial and telomeric rRNA loci (Figures 4.19 & 4.22). These observations are in concordance with findings from budding yeast **(Chan *et al.*, 2014; El Hage *et al.*, 2014; Wahba *et al.*, 2016)**, fission yeast **(Hartono *et al.*, 2018; Legros *et al.*, 2014)**, and *Arabidopsis* **(Xu *et al.*, 2017)** which identified the above regions as hot spots for R-loop formation. Importantly, signals over rRNA genes are still the strongest even after accounting for the repetitive nature of these genes. Consistent with an established and well known correlation between the rate of transcription and R-loop formation in mostly all systems, and similar to the pattern recognized for *S. cerevisiae* **(Chan *et al.*, 2014; El Hage *et al.*, 2014; Wahba *et al.*, 2016)**, *S. pombe* **(Hartono *et al.*, 2018)** and *A. thaliana* **(Xu *et al.*, 2017)**, signals were strongly enriched over highly transcribed protein-coding RNA Pol II-transcribed genes and tf2 retrotransposons genes. Of the most highly enriched protein coding genes are; heat shock protein (hsp) genes (e.g., hsp90); genes involved in ribosome biogenesis and protein synthesis such as 40S (e.g., rps1071) and 60S (e.g., rpl1803) ribosomal protein genes; and translation elongation factors (e.g., tef3 and tef103) genes (Figure 4.13). These findings are consistent with previous R-loop mapping studies on budding yeast **(El Hage *et al.*, 2014; Wahba *et al.*, 2016)** and fission yeast **(Hartono *et al.*, 2018)** which identified these regions as hot spots for R-loop formation, especially in RNase H-lacking cells. All the methods captured very strong signals over mitochondrial genome in agreement with previous findings **(El Hage *et al.*, 2014)** and consistent with a role of R-loops in mitochondrial DNA replicaiton **(Holt, 2019)**

This strong agreement with previous reports strongly confirm that signals detected by my methods represent genuine R-loops. All the above observations have been revealed by all my methods, the dsChIP-exo, ssChIP-exo and SMART-DRIPc for the untreated and RNIII/P1-treated samples with some differences between WT and *rnh1Δ rnh201Δ* cells. Detected signals exhibit features previoulsy established from different organisms (including fission yeast) for real R-loops **(Hartono *et al.*, 2018; Sanz *et al.*, 2016; Xu *et al.*, 2017)**. They are codirectional

with transcription, map to transcription units and form over gene bodies, and most of signals are RNase H sensitive. Only in RNase H-depleted cells, signals extend beyond gene bodies, but at lower levels compared to the main genic signal. These extragenic signals aren't detected in WT cells, because they are RNase sensitive (degraded by RNase H) (Figure 4.13, signals over tRNAs, snu3 and sno13). This is a clear distinction for real R-loops from false signals that may be detected over some intergenic regions in WT, but don't respond to RNase H deletion.

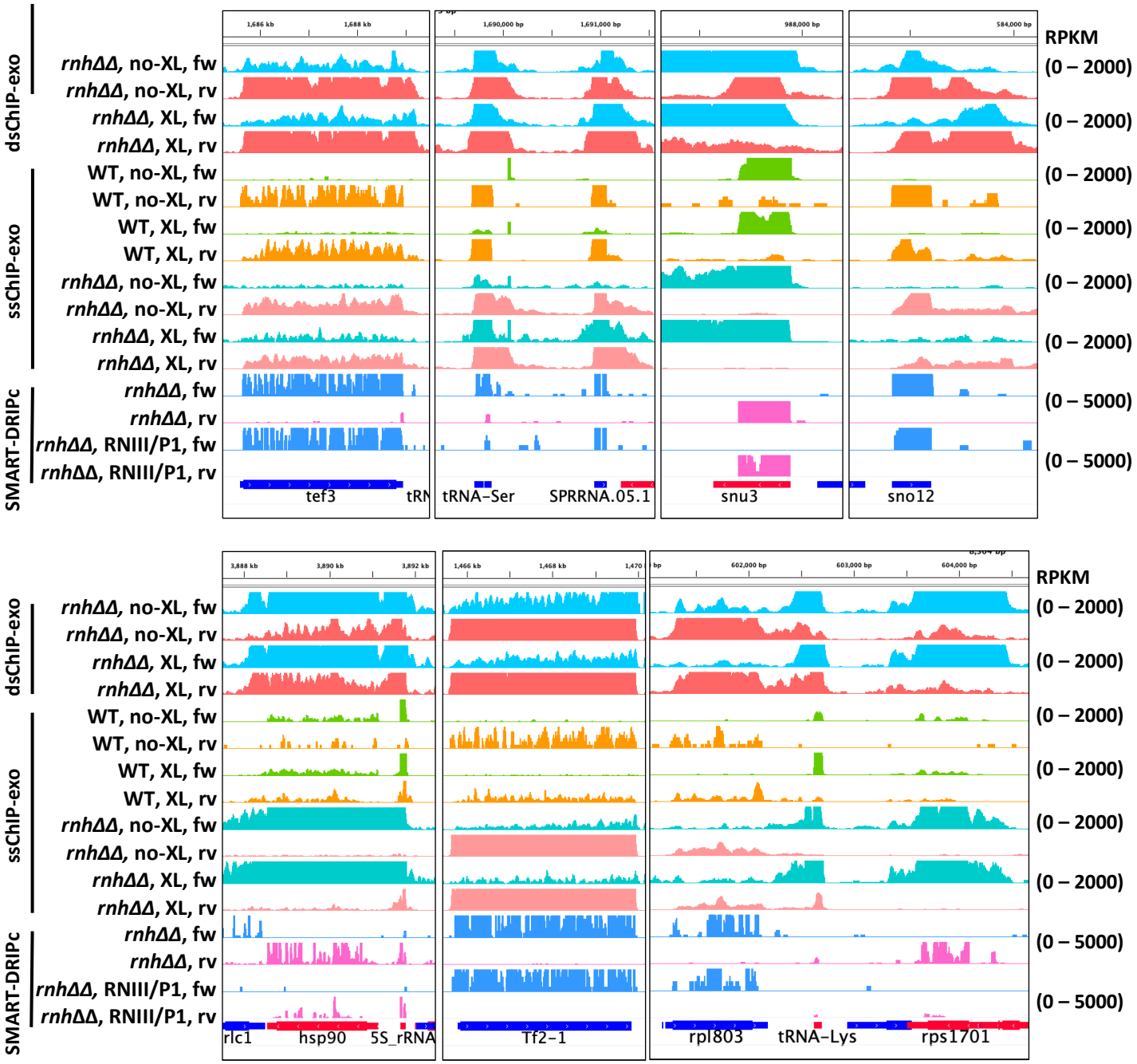



**Figure 4.13. Representative examples for signal-enriched regions matching known R-loop hot spots.** Snap shot of genome browser viewer showing signals forming over regions known as hot spot for R-loop formation. Also note that signals mapped by different methods, over enriched regions, show variability in directionality and in signal strengths over non-template strand.

### 4.3.9 No correlation exists between fw-strand and rv-strand R-loop signals in WT cells

After confirming the specificity of the signals, I moved to study the reproducibility of correlation exhibited by previous external ChIP-exo data (Figure 4.5) for forward (fw)- and reverse (rv)-strand signals, in order to verify the double R-loop formation model. Hereafter, whenever talking about R-loop signal existence in relation to transcription direction of a specific gene, I will call the signals as either template-strand or non-template-strand signals, referring to transcriptionally co-directional or transcriptionally anti-directional signals, respectively. For general existence of signals, regardless of transcription direction, I will call signals as forward-strand or reverse-strand signals for signals forming over plus and minus DNA strands, respectively.

Metagene analysis was performed to understand the global correlation between forward- and reverse-strand signals captured by different methods over enriched regions. For this analysis, forward strand signals were first ranked in a descending order based on signal strength. Next, reverse-strand signals were aligned to corresponding forward ones. External DRIP data from Martienssen lab (**Castel *et al*., 2014)** was used as an example for non-directional and non-strand specific data and as a control to see how the distribution of metagenic signal should look for non-directional methods. Explicitly, metagene analysis for external DRIP showed absolute coincidence between both strand signals that, obviously, both strand signals carbon-copied each other, meaning that they looked identical in peaks heights and distributions. To some extent, this pattern was observed for *rnh1Δ rnh201Δ* cells dsChIP-exo forward-strand and reverse-strand metagenic signals which had similar strengths and overlapping central peaks (Figure 4.14, see metaplots).

Knowing that ChIP-exo provides strand-specific information for proteins, these results from dsChIP-exo for *rnh1Δ rnh201Δ* cells, similar to external ChIP-exo in Figure 4.5, suggested that for every signal forming on one DNA strand, an overlapping signal forms on the other strand. By another meaning, both DNA strands form R-loop signals at the same frequency and over the same sequence. These observations, initially, suggested the formation of double R-loops almost over all enriched regions. Differently, the pattern and correlation revealed by external ChIP-exo and my dsChIP-exo, wasn't generated by ssChIP-exo for the same *rnh1Δ rnh201Δ* cells, which started to show difference in signal strengths and peaks distribution (Figures 4.14, compare position of peaks and scaling of metaplots).

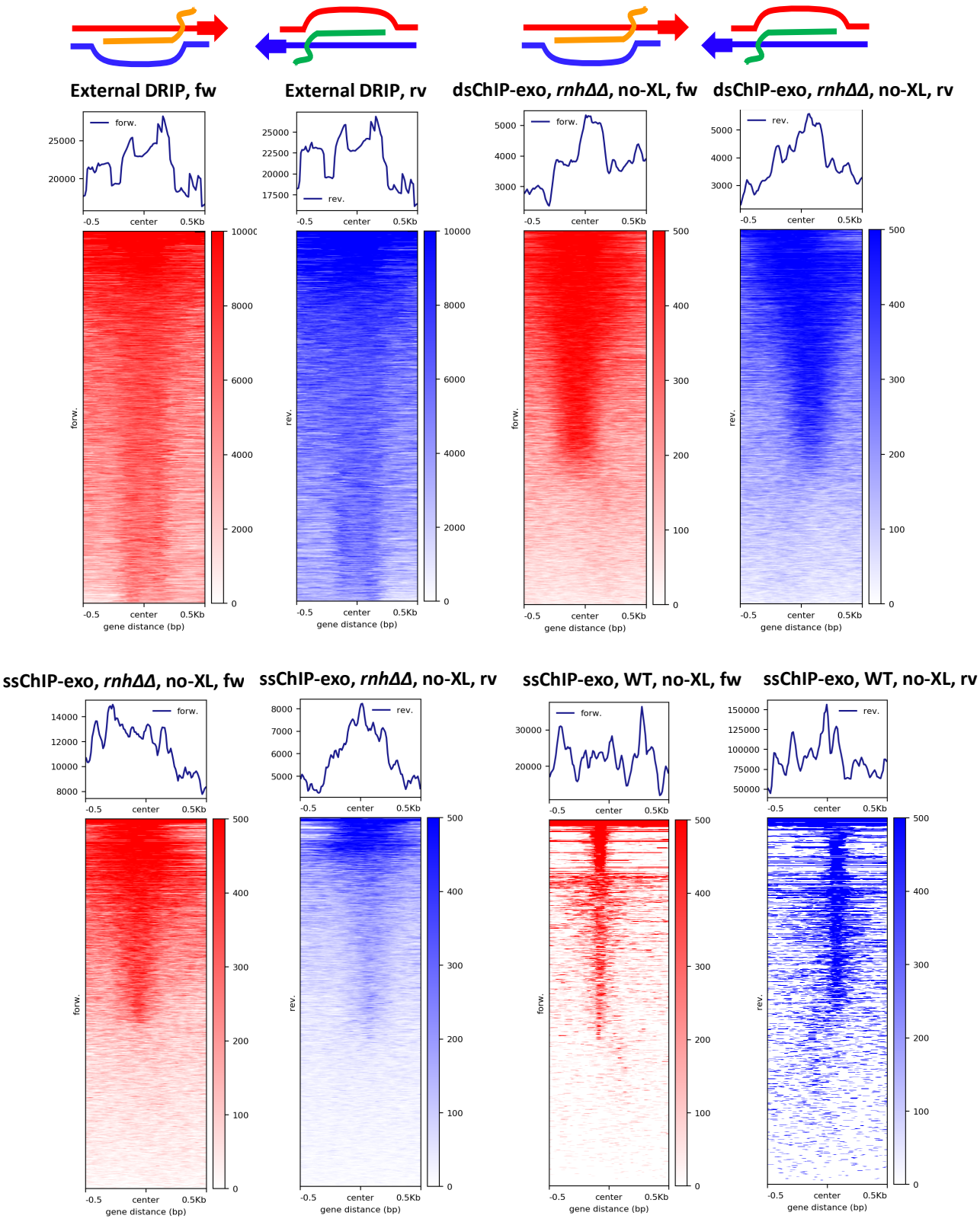



**Figure 4.14. Metagenic analyses for forward- and reverse-strand R-loop signals of different methods.** Metaplots and heat maps for R-loops signals distribution over both DNA strands in 1 Kb window, showing strong correlation between forward (fw)- and reverse (rv)-strand R-loop signals mapped by external DRIP and my dsChIP-exo, but not ssChIP-exo. External DRIP data was used as a control for positive correlation. Schematic representation for mapped R-loop-forming DNA strand is shown.

Importantly, the correlation/pattern was disrupted for the WT ssChIP-exo signals where the reverse-strand metagenic signal was five-fold stronger than the forward-strand signal. Moreover, it was hard to define a clear overlapping in peaks (Figure 4.14). Strikingly, SMART-DRIPc for *rnh1Δ rnh201Δ* cells showed no correlation at all. There was neither overlap in peaks nor concordance in their strengths. Instead, they revealed variable peak distribution and strengths for forward compared to revere signals (Figure 4.15).

Overall, these observations suggest that the correlation between forward- and reverse-strand R-loops signals isn't reproducible among different methods and don't support genome-wide double R-loop formation. Moreover, this refers to a difference in the correlation for RNase H-depleted and WT cells where it tends to disappear in the later. Finally, the difference in the degree of correlation among methods, refers to a difference in their directionality.

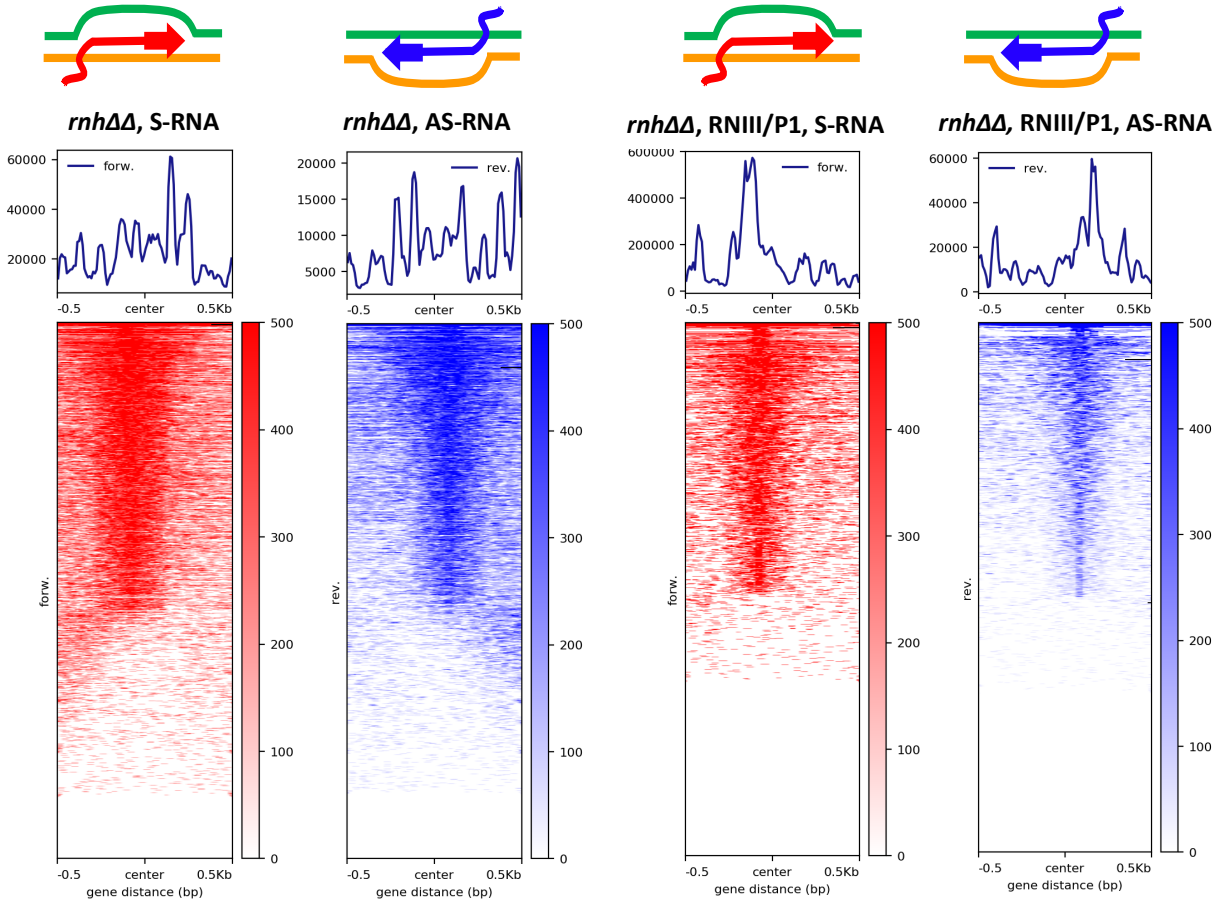



**Figure 4.15. Metagenic analyses for forward and reverse R-loop signals of SMART-DRIPc.** Metaplots and heat maps for distribution of R-loops signals in 1 Kb window, showing absence of correlation between forward and reverse R-loop signals of SMART-DRIPc. Schematic representation for mapped R-loop-forming RNA strand is shown. Note that, different form DNA-based methods, fw and rv signals here represent sense RNA (S-RNA) and antisense RNA (AS-RNA) strands, respectivley.

### 4.3.10 ssChIP-exo and SMART-DRIPc show unidirectional R-loop signals genome wide

Next, I studied directionality of methods and tested double R-loop formation by directly examining individual R-loop signals generated by different methods. In fact, during my investigation for R-loop-enriched regions, I noticed a difference in R-loop profiles regarding directionality of signals captured by my different methods. As a rule, and consistent with established facts about R-loops formation, all my methods detected signals, mainly, forming over the template DNA strand and oriented toward transcription direction (Figure 4.13). This means for DNA-based methods (dsChIP-exo and ssChIP-exo), the signal maps to the reverse (minus) strand when the gene is transcribed in the sense direction, and exists on the forward (plus) strand when the gene is in the antisense direction whereas for the RNA-based SMART-DRIPc, the signal maps to the forward (plus) strand in case of sense genes, and to the reverse (minus) strand in case of antisense genes.

Remarkably, and regardless of the consensus regarding the formation of signals over template strand, clear differences existed among the different methods regarding signal strength on non-template strand. Similar to the previous external ChIP-exo results for RNase H-depleted fission yeast cells **(Ohle *et al.*, 2016)**, my dsChIP-exo detected relatively strong signals on the non-template strand. Explicitly, short genes such as tRNAs, 5s-rRNAs, and telomeric and mitochondrial rDNA repeats showed the strongest non-template relative to template strand signals (Figure 4.13). Based on these observations, I selected these regions for studying spearman correlation for forward and reverse signals generated by different methods. I used the non-directional external DRIP as a control for a very strong degree of correlation. dsChIP-exo revealed a weak to mild degree of correlation over tDNA and rDNA, respectively (Figure 4.16). Surprisingly, mitochondrial tDNA and telomeric and mitochondrial rDNA showed the strongest correlation. Explicitly the non-template strand signal tended to be as strong as that of template strand, referring to double R-loop formation over these regions (see final discussion).

Surprisingly, ssChIP-exo detected weaker signals on non-template strand relative to those on template strand and compared to dsChIP-exo. This was very clear for both of the WT and RNase H-depleted cells (Figures 4.13, 4.18 and 4.19). These observations suggest that this method captures only the DNA strand of the hybrid and confirm the enhanced directionality of ssDNA-based compared to dsDNA-based methods, in agreement with the ssDRIP **(Xu *et al.*,**

**2017)**. Compared to dsChIP-exo, ssChIP-exo showed weaker correlation over tRNAs, 5s-rRNAs, and rRNA genes. In fact, ssChIP-exo revealed an anticorrelation for forward and reverse strand signals in WT cells (Figures 4.24 & 4.25). Regardless of the correlation, these R-loops still have high levels on both DNA strands which doesn't exclude double R-loop formation at these genes (Figure 4.13, 4.24 & 4.25).

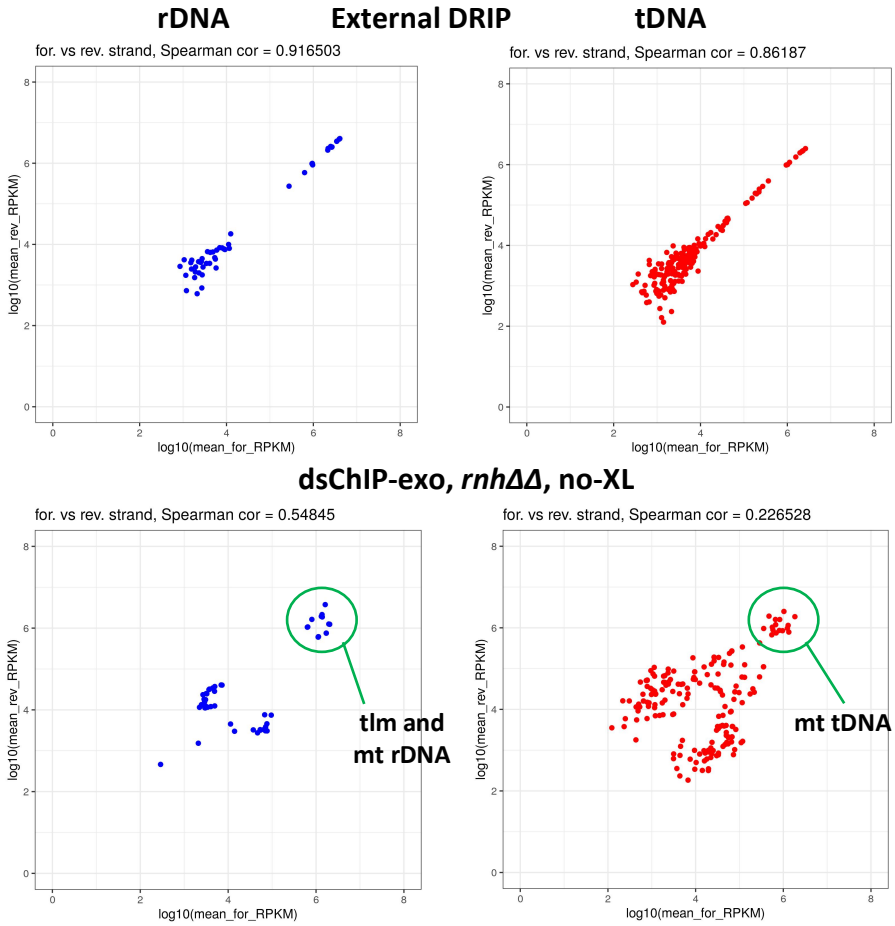



**Figure 4.16. Spearman correlation analysis for dsChIP-exo data for forward- and reverse-strand R-loop signals over rRNA and tRNA genes.** Analysis shows mild correlation over rDNA (blue) but weak correlation over tDNA (red) in general. Circled points refer to telomeric (tlm) and mitochondrial (mt) genes of highest correlation and highest R-loops levels. An external DRIPc was used as a control for strong correlation.

Strikingly, SMART-DRIPc for RNase H-depleted cells produced signals that are strictly stranded which means they map only to the transcription direction (Figure 4.13). SMART-DRIPc revealed a mild to strong anticorrelation between forward and reverse R-loops signals detected for tRNA and rRNA genes (Figure 4.17). Surprisingly, this pattern was similar for the untreated and RNIII/P1-treated samples suggesting that RNase III isn't required for directional mapping and that the SMART-DRIPc is inherently directional under the experimental

conditions used. Detection of bidirectional signals over tRNA and rRNA genes using dsChIP-exo and ssChIP-exo, but not SMART-DRIPc is reminiscent with a similar situation where the ssDNA-based R-ChIP **(L. Chen *et al.*, 2017)**, but not the RNA-based RR-ChIP **(Tan-Wong *et al.*, 2019)** detected sense and anti-sense signals over these regions in mammalian cells.

Overall, ssChIP-exo and SMART-DRIPc results are consistent with directional R-loop profiling studies on *S. pombe* using DRIPc **(Hartono *et al.*, 2018)**, *Arabidopsis* using ssDRIP **(Xu *et al.*, 2017)**, and mammalian cells using DRIPc, R-ChIP and RR-ChIP **(L. Chen *et al.*, 2017; Sanz *et al.*, 2016; Tan-Wong *et al.*, 2019)** which detected few regions with sense-antisense R-loops. These studies established and confirmed the formation of unidirectional R-loops oriented with transcription direction. Importantly, the dominance of unidirectional signals produced by both ssChIP-exo and SMART-DRIPc doesn't support genome-wide formation of double R-loops. Nevertheless, ssChIP-exo, and to a lesser extent, SMART-DRIPc still show R-loops over both template and non-template strands of some rRNA and tRNA genes.

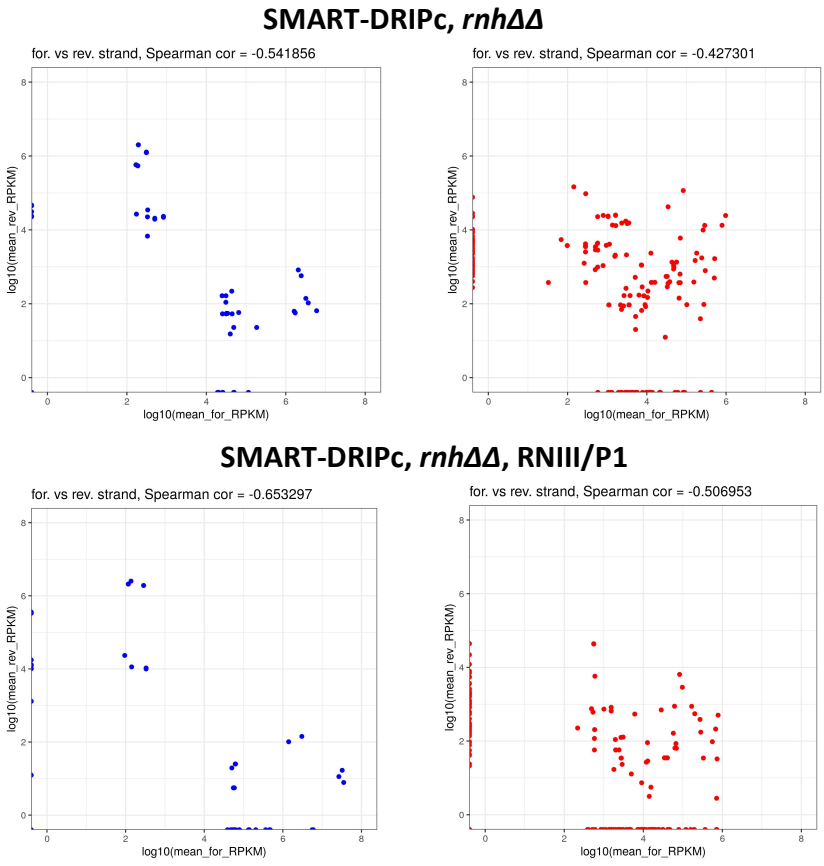



**Figure 4.17. Spearman correlation analysis for SMART-DRIPc data for forward and reverse R-loop signals over rRNA and tRNA genes.** Analysis shows mild to strong anti-correlation between forward and reverse R-loops signals of SMART-DRIPc RNA forming over rDNA (blue) and tDNA (red).

Although this will not change the final conclusion, there were some differences, at the technical level, between SMART-DRIPc and external DRIPc **(Hartono *et al.*, 2018)**. It has been shown that the RNA-based DRIPc captures dsRNA signals that required treatment with RNase III in order to deliver stranded signals mapping to the template DNA **(Hartono *et al.*, 2018)**. The same authors detected double signals belonging to dsRNA at the dg/dh pericentromeric repeats which were eliminated by RNase III and found to be sensitive to the exosome, but not endogenous RNase H expression. Surprisingly, and despite the great sensitivity of the SMART-DRIPc, I didn't detect such dsRNA signals even without prior RNase III treatment.

### 4.3.11 Most of R-loop signals are sensitive to RNase H in WT cells

Practically, for an R-loop to contribute to dsRNA formation, the RNA strand of the hybrid has to be resistant or stabilized, at least partially, against the different cellular nucleases. Of these nucleases are RNase H enzymes, the main hybrid-resolving and conserved enzymes in all cellular organisms. It's hard to imagine an unstable and actively degraded R-loop to participate in dsRNA formation. To examine the stability of R-loops against RNase H, I compared the R-loop profiles of WT yeast cells (with physiological levels of RNase H) to those of the RNase H-depleted *rnh1*Δ *rnh201*Δ cells (which accumulate R-loops). For this, I used the ssChIP-exo data generated for non-crosslinked cell types to avoid variations potentially induced by fixation. As shown in Figure 4.18, ssChIP-exo detected only 749 signal peaks for the non-crosslinked WT compared to 1994 signal peaks for the non-crosslinked *rnh1*Δ *rnh201*Δ cells, referring to instability of R-loops forming over ~60 to 70% of R-loop-forming loci against physiological levels of RNase H. These observations confirm previous findings form *S. pombe* **(Hartono *et al.*, 2018)** and *S. cerevisiae* **(Wahba *et al.*, 2016)** that WT cells accumulate less R-loops. Obviously, signals were generally depleted genome-wide in WT compared to *rnh1*Δ *rnh201*Δ cells, implying global instability of R-loops against RNase H (Figure 4.18).

Strikingly, signals over heat shock proteins (hsp) genes were the most sensitive and almost abrogated in WT. Consistent with a role of RNase H in retro-transposition in yeast and in agreement with findings from *S. cerevisiae* using S1-DRIP-seq **(Wahba *et al.*, 2016)**, signals over tf2 retrotransposons were also depleted in WT cells. Besides, signals detected over a

subset of actively transcribed genes such as ribosomal-proteins coding genes (e.g., rps7 and rpl402) and genes linked to translation, were among the sensitive category (Figure 4.18).


**Figure 4.18. The majority of R-loop signals are sensitive to RNase H. (A)** Venn diagram analysis showing number and overlap of R-loop signals detected by ssChIP-exo for WT and *rnh1Δ rnh201Δ* cells. **(B)** snap shot of genome browser showing examples for R-loops signals decreased in WT compared to *rnh1Δ rnh201Δ cells*.

These observations suggest the instability of R-loops forming over these regions. Given these findings, it's unlikely that these unstable R-loops would participate in structural features such as dsRNA. Importantly, all the above sensitive R-loop signals had weak to mild strengths. An exception for this condition is the mitochondrial R-loop signals which were also weaker in WT compared to the *rnh1*Δ *rnh201*Δ cells (Figures 4.21, 4.24 & 4.25). Despite being sensitive, mitochondrial R-loop signals still have high density on both DNA strands (Figures 4.21, 4.24 & 4.25). Of course, it should be considered that mitochondrial genome consists of DNA sequence repeats that may account for this strong signal enrichment.

Notably, some of the weak R-loop-forming regions with antisense transcription (R-loop signal forming on forward strand) show weak R-loop signals on reverse strand in WT cells, which might suggest double R-loop formation (Figure 4.18B, extensive genomic region). However, this observation should be considerd carefully as the spiky signal on reverse strand seem to represent a false background signal, but not real R-loops. This explanation is based on the fact that these signals extend over intergenic regions in WT. Besides, these extensions don't increase by RNase H deletion as expected for intergenic signals (as seen in Figure 4.13), i.e., as a rule, real extragenic signals decrease in WT, but increase with RNase H deletion.

Remarkably, R-loop signals over a subset of genes weren't weaker in WT compared to *rnh1*Δ *rnh201*Δ cells, implying stability of R-loops over these regions against RNase H. These included signals of some 40S and 60S ribosomal proteins genes, and translation elongation genes (e.g., tef3 and tef102). These regions are characterized by very strong R-loop signals (Figure 4.19). From the stability perspective, this R-loop category is the best candidate that can be envisaged to have the capacity to form dsRNA, if possible. However, these stable R-loops form only in one direction. For this reason, it's unlikely that these regions would form double R-loops or dsRNA in a double R-loop dependent manner. Instead, dsRNA formation through interaction between free antisense RNA and these single R-loops may be more plausible. Strikingly, similar to telomeric rRNA (18S and 28S rRNA) genes known to accumulate RNase H-insensitive R-loops, tRNA and 5s-RNA genes showed very strong R-loop signals that were similar in strength or slightly weaker in WT cells compared to *rnh1*Δ *rnh2*Δ cells, supporting the partial stability of R-loops over these genes. Similar to mitochondrial DNA, telomeric rDNA formed R-loops on both sides, making them possible candidates for double R-loop formation. Importantly, non-template-strand signals of 5s-rRNA and tRNA genes, at the periphery of heterochromatin, were sensitive and decreased in WT (Figure 4.19).

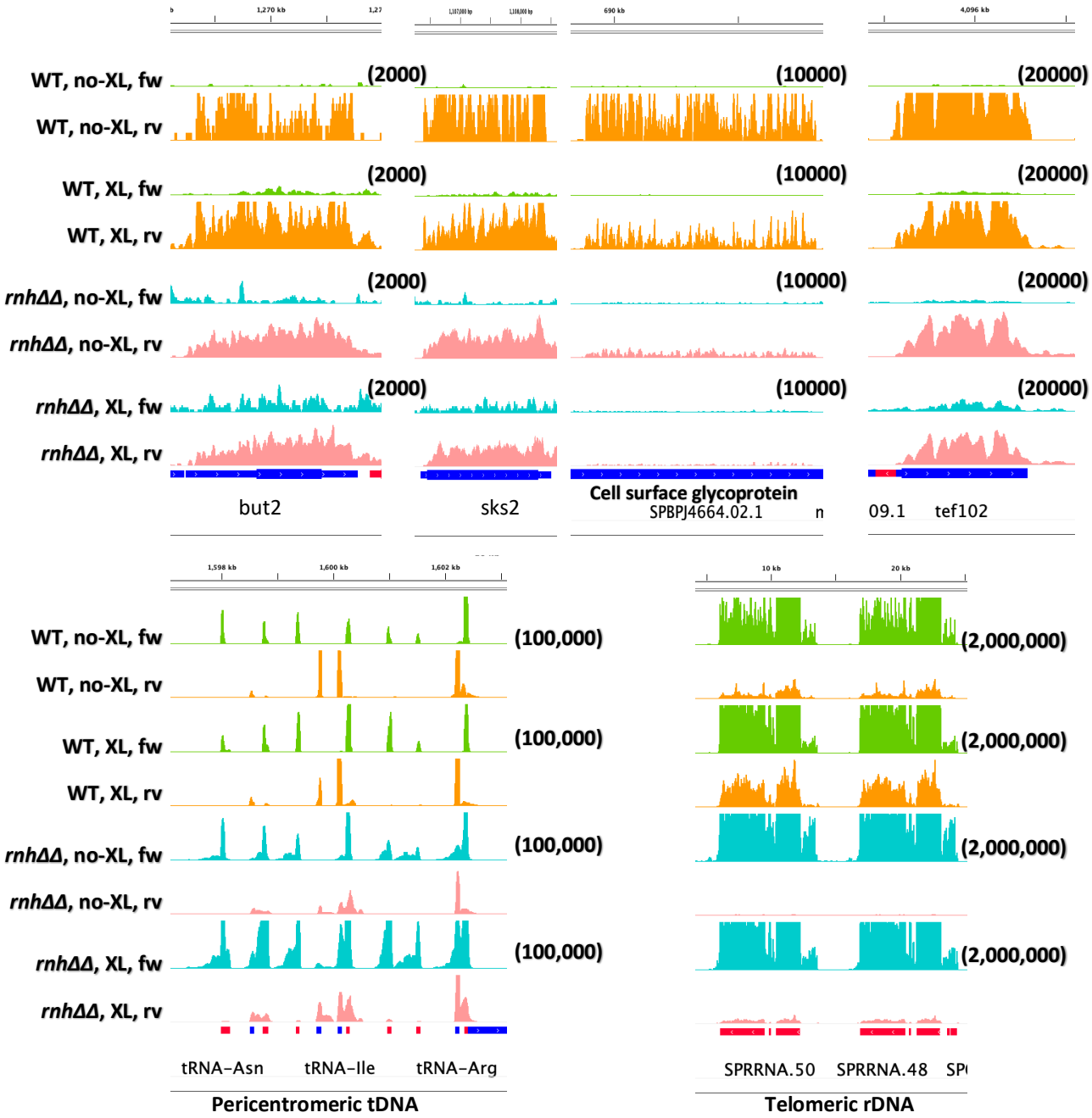



**Figure 4.19. Representative regions of RNase H-insensitive R-loop signals.** Snap shot of genome browser showing regions with R-loops signals not decreased in WT compared to *rnh1 rnh201 cells*.

My observations support a recent study which identified two classes of R-loops signals, sensitive and insensitive R-loops signals, based on sensitivity to endogenous RNase H in fission yeast (**Hartono *et al.*, 2018**). However, this study didn't support formation of R-loops over tDNA regions in WT cells, identifying tDNA R-loops among the sensitive category. In contrary, I detected very strong signals over these regions in WT cells. The reason for this discrepancy is unclear, but one possibility is that R-loops over these regions in WT cells are

very short that they can be easily denatured during phenol chloroform extraction for the DRIPc which isn't applicable for SMART-DRIPc. Indeed, ssChIP-exo revealed that WT R-loops signals over tRNA genes are affected mainly in width but not in strength. More precisely, at tRNAs and 5s-rRNAs genes, in the *rnh1*Δ *rnh2*Δ cells, the signals extend in width outside the gene body, but do not increase in height (Figures 4.13 & 4.19).

Identifying sequence and chromatin features of weak and strong R-loops captured using the current methods, may be helpful to distinguish these R-loop categories. In fact, R-loop formation was found to be associated with GC content and GC skew in *Arabidopsis* **(Xu *et al.*, 2017)** and mammalian cells **(Ginno *et al.*, 2013; Ginno *et al.*, 2012; Sanz *et al.*, 2016)**. Also, RNase H insensitivy has been linked to high GC content and high positive GC skew in mammalian cells **(Crossley *et al.*, 2020)**. Similarly, in fission yeast, RNase H insensitive R-loop category was found to be significantly associated GC skew **(Hartono *et al.*, 2018)**. Moreover, R-loop formation in fission yeast was found to be significantly associated with active transcription and transcriptiton elongation chromatin factors, similar to findings from to *Arabidopsis* and mammalian cells **(Sanz *et al.*, 2016; Xu *et al.*, 2017)**.

### 4.3.12 RNase H-insensitive R-loop signals overlap with asRNA and dsRNA signals

While most of R-loop signals showed sensitivity to RNase H, R-loops signals over some regions weren't eliminated in WT compared to *rnh1*Δ *rnh2*Δ cells, showing partial or complete insensitivity to RNase H. However, these R-loops signals formed only on one direction. If R-loops of these regions were able to form dsRNA, this wouldn't be through double R-loops but rather through single R-loops and free asRNA. Supporting this possibility, I found that these R-loops are associated with regions of antisense transcription (Figure 4.20). For further support and confirmation, I used recent mature RNA and nascent RNA sequencing (RNA-seq and NET-seq, respectively) public data for WT fission yeast cells from Morillon lab **(Wery, Gautier, Descrimes, Yoda, Vennin-Rendos*, et al.*, 2018)**. Surprisingly, I found that some of these insensitive R-loop signals overlap with mature and nascent antisense RNA signals (Figure 4.20). In general, these observations are in line with previous findings that R-loops form over regions of sense-antisense transcription in budding yeast **(Chan *et al.*, 2014)**.

Next, I sought to check whether the same R-loops signals are associated with dsRNA signals. Due to the lack of dsRNA-seq data for WT fission yeast, I used fission yeast DRIPc data that was reported to detect dsRNA signals at over 35% of the genome **(Hartono *et al.*,**

**2018)**. Surprisingly, these RNase H-insensitive R-loops signals overlapped with dsRNA signals. Notably, however, checking the nascent and mature RNA profiles of different data showed that some of these dsRNA signal loci exhibit only single-direction transcription (Figure 4.20). It is hard to imagine or explain how dsRNAs can form over those regions. Importantly, telomeric rDNA (Figure 4.20) and mitochondrial DNA (data not shown), but not tDNA (Figure S3) regions were associated with antisense transcription and dsRNA formation, suggesting that mitochondrial genes may be the best candidates to be investigated for double R-loop-dependent dsRNA formation.

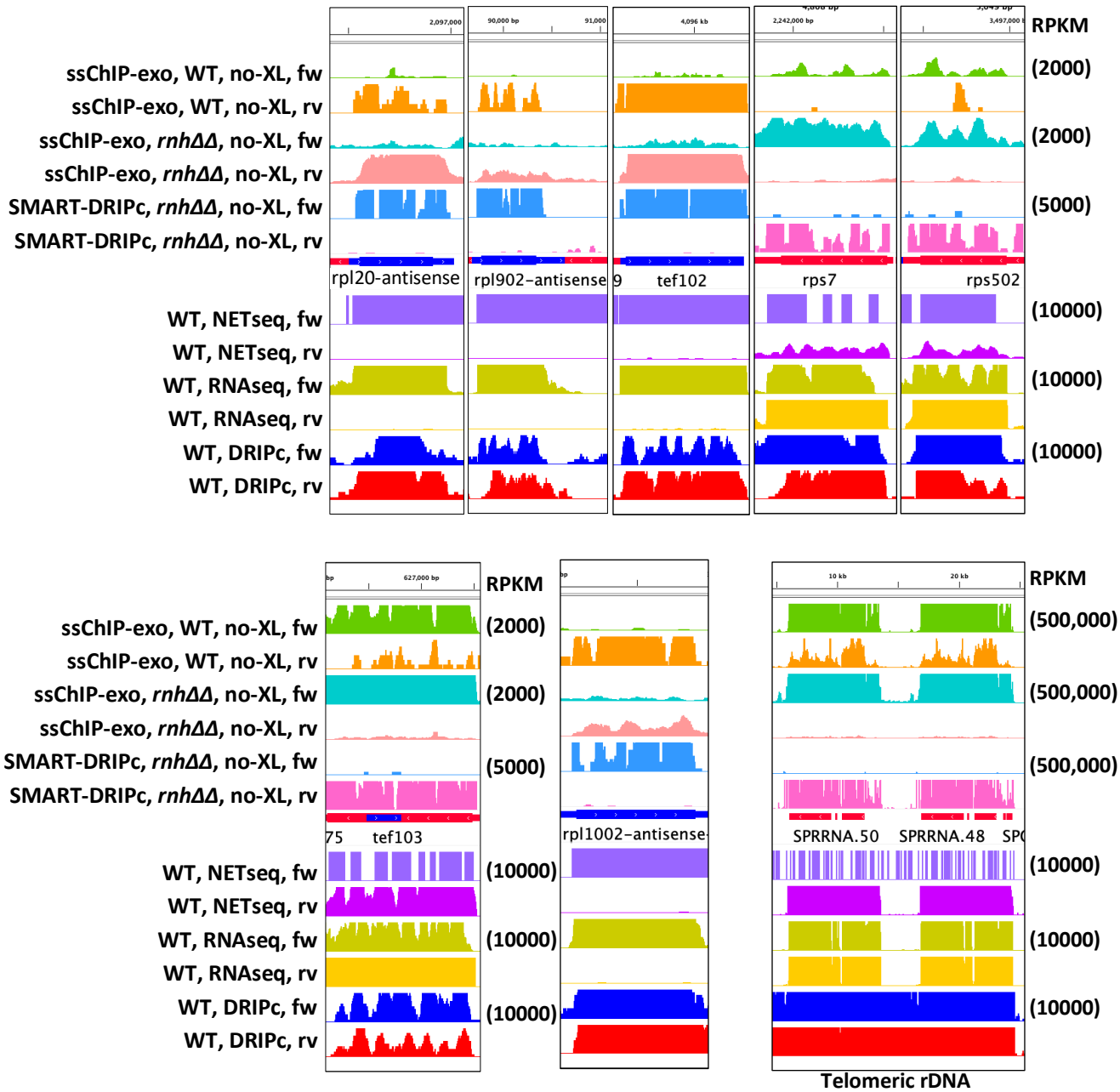



**Figure 4.20. RNase H-insensitive R-loop signals coincide with regions of AS transcription and dsRNA formation**. Snap shot of genome browser showing R-loop signals overlapping with mature RNA, nascent RNA and dsRNA signals

### 4.3.13 RNase H-insensitive R-loop signals form over reverse DNA strand

Excitingly, I detected a difference in RNase-H-sensitivity of R-loop signals based on their formation direction (orientation). More precisely, forward-strand R-loop signals were more increased by RNase H deletion compared to reverse-strand signals, suggesting that forward-strand R-loops are more sensitive to RNase H in WT cells. Intriguingly, RNase H-insensitive R-loop signals tended to form on the reverse DNA strand and over sense-transcribed genes (genes transcribed in the normal sense transcription direction). This is exactly the case for the above ribosome biogenesis and protein synthesis genes (but2, sks2, tef102) with stable R-loops (Figures 4.19 & 4.22).

These observations suggest that R-loop orientation (hybrid formation direction) modulate its sensitivity to RNase H. Strikingly, RNase H deletion increased mitochondrial forward-strand R-loop signals ten times compared to those of WT, but slightly affected those of reverse strand. This confirms that forward strand R-loops are more sensitive to RNase H compared to reverse strand signals, while the opposite may be true for reverse-strand compared to forward-strand signals. This effect was clearly shown over genome browser and using spearman correlation analyses (Figures 4.21, 4.24 & 4.25).

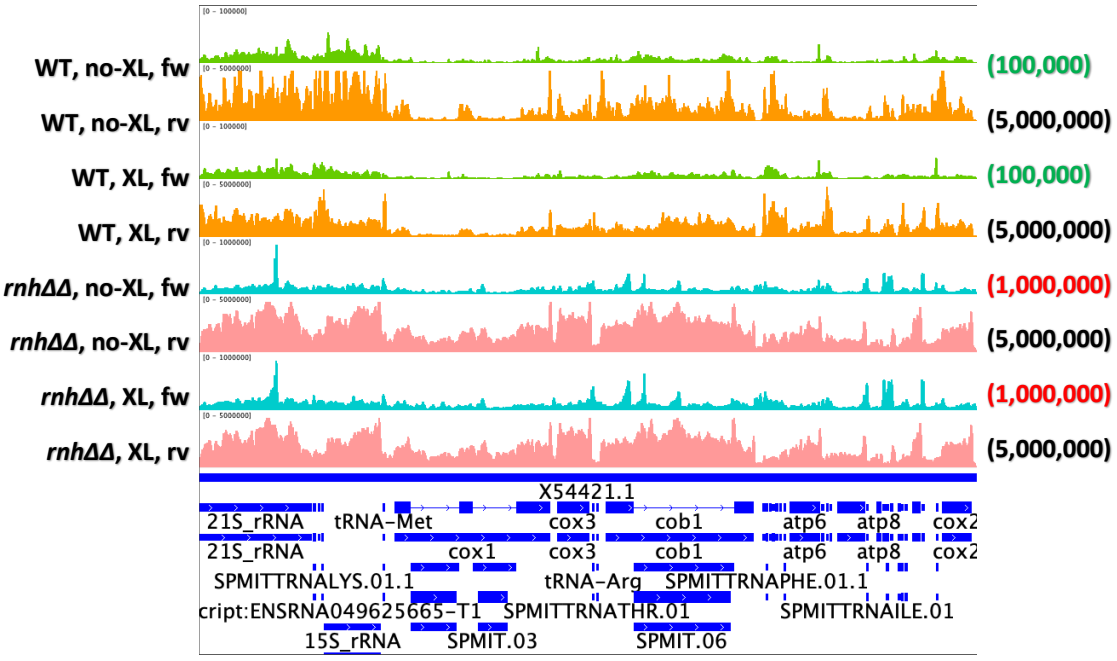



**Figure 4.21. RNase H deletion strongly increased forward-strand but not reverse-strand R-loop signals over mitochondria.** Snap shot of genome browser showing R-loop profiles over forward and reverse strand of mitochondrial DNA. Notice scaling.

### 4.3.14 RNase H deletion may induce global transcriptomic and proteomic changes and affect cellular activities

At first glance, R-loop signals over most of forming regions showed sensitivity to RNase H that these signals were only detected in RNase H depleted cells. Further inspection revealed strong R-loop signals forming over template strand (reverse DNA strand) of some sense-transcribed genes in WT cells. Excitingly, these signals got decreased in strength or even depleted with RNase H deletion (Figures 4.19 & 4.22). Since R-loops formation is induced mainly by transcription, this observation suggests a change in transcription over these genes, which may have affected R-loop formation. By different words, decrease in R-loop levels over specific genes in *rnh1Δ rnh201Δ* cells, suggests that these genes have lower transcription rate compared to WT cells. Intriguingly, most of these genes are related to ribosome biogenesis and protein synthesis. Notably, cell surface glycoprotein genes (SPAPB2C8.01.1 and SPBPJ4664.02.1) and DNA damage checkpoint rad24 gene exhibited the same effect. This implies that prolonged manipulation of RNase H and R-loop levels may affect vital cellular processes and transcriptional and translational profiles. Notably, this effect may be independent of local enzymatic activity of RNase H at these regions.

Further inspection showed that the decrease of R-loop signals over template strand of these regions was accompanied by an increase of R-loop signals over non-template strand in *rnh1Δ rnh201Δ* compared to WT cells (Figures 4.19 & 4.22). Although this refers to a possible regulatory role mediated by RNase H, it's hard to conclude whether this increase in R-loop levels over non-template strand matches an increase in antisense transcription. In general, this observation is in line with an anticorrelation between sense and antisense transcription and between forward-strand and reverse-strand R-loops.

It sounds complicated to decipher this phenomenon as it can be explained by either an effect for global R-loop accumulation as a result of RNase H deletion or by an impact for deficiency of RNase H in the cell. Global accumulation of RNase H sensitive R-loops on non-template (forward) strand may have impeded sense transcription and accordingly decreased R-loop formation on template (reverse) strand. On the other hand, deficiency in cellular levels of RNase H may have perturbed sense transcription, especially for protein coding genes. This may have increased oppourtunistic antisense transcription and R-loop formation on forward strand. Anway, the final outocome would be a decrease in gene expression for vital genes

involved in multiple celluar processes, especially translation and protein synthesis. It's known that R-loop formation and levels are dependent on transcription rate. However, the level of R-loops in this situation can't be used to infer transcription rate. It's hard to make a connection here in the presecence of two opposing responses and a third variable which is RNase H. That's why mature and nascent transcriptome profilings are required to understand this phenomenon.

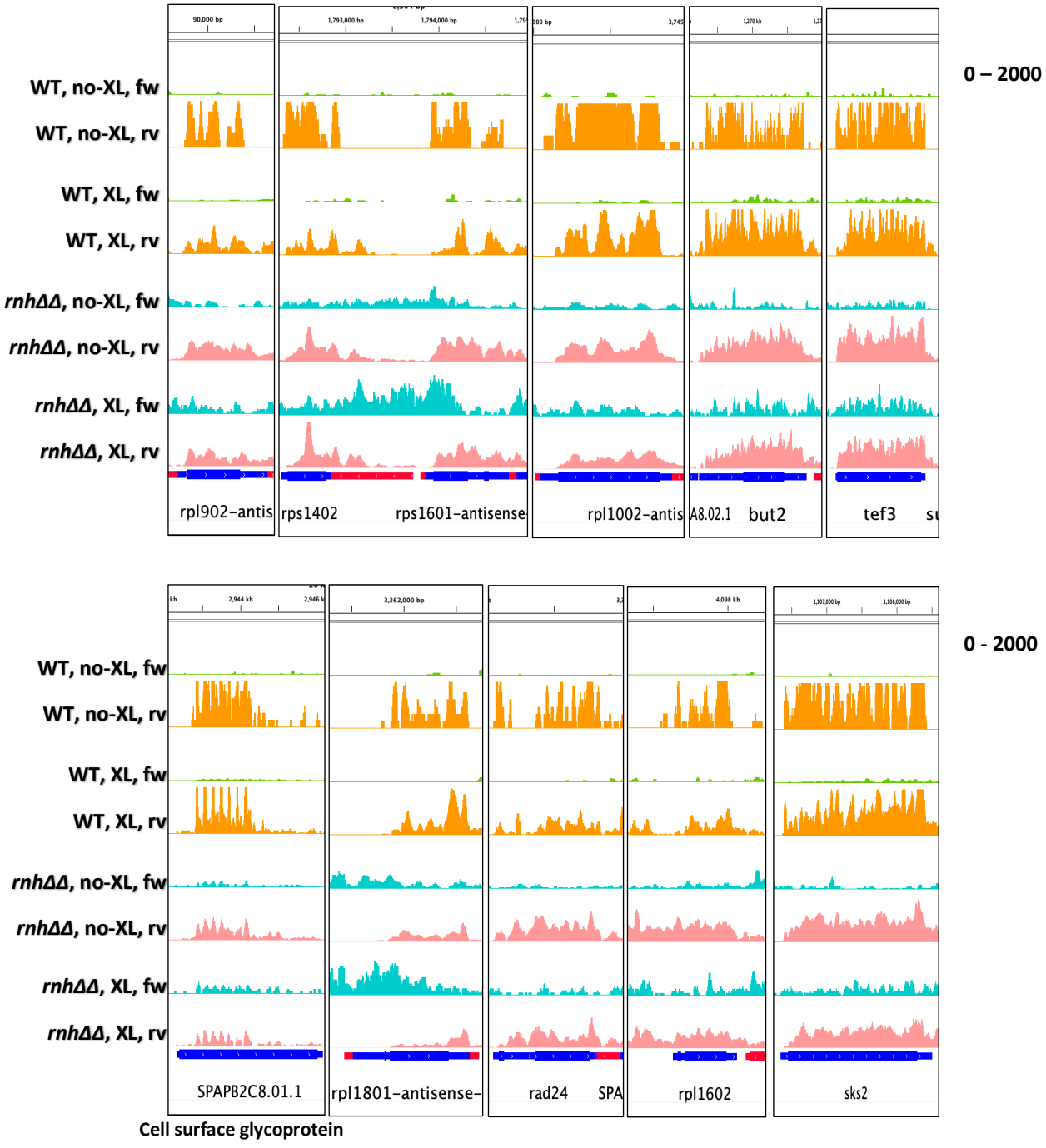



**Figure 4.22. RNase H deletion decreased R-loop signals over template strand of ribosome biogenesis genes.** Snap shots of genome browser, showing R-loop profiles over termplate and non-template strand of sense ribosome biogenesis and protein synthesis genes.

### 4.3.15 RNase H deletion increased fw-strand while decreased rv-strand R-loop signals

In ultimate contrast and different from the above sense genes, RNase H deletion increased R-loop signals over template strand while decreased signals over non-template strand of all antisense transcribed genes. These observations suggest distinct impacts, for RNase H deletion, on gene expression based on transcription direction of R-loop forming genes. Superficially, these observations are consistent with a previous finding that manipulation of RNase H levels (either RNase H deletion or overexpression) may impact the transcriptome of fission yeast cells **(Hartono *et al.*, 2018)**. Thorough examination to understand the reason behind these opposing effects for RNase H deletion on R-loop signals of both sense and antisense genes, showed that RNase deletion induced genome-wide and contrasting effects on R-loop signals of both DNA strands independent form transcription direction. In essence, RNase H deletion increased R-loop signals over forward strand and, meanwhile, decreased signals over reverse strand compared to WT cells (Figures 4.18, 4.19, 4.22 and 4.23).

This effect was also shown for telomeric rDNA repeats (Figures 4.19 & 4.20). Moreover, it was clearly revealed by spearman correlation analysis for forward and reverse strand R-loop signals over tDNA and rDNA for WT and *rnh1Δ rnh201Δ* cells (Figures 4.24 & 4.25). For rRNA and tRNA genes, spearman correlation analyses showed that this effect is more apparent for sense genes where R-loop signals form on the reverse strand and tends to be weak on forward strand. RNase H deletion showed a clear increase of forward strand signals and decrease of reverse strand ones. On the other hand, R-loops of antisense genes form on the forward strand but not on the reverse one. RNase H deletion showed clear increase in forward strand signals, while the decrease in reverse strand signal wasn't obvious as this strand has low levels any way (Figure 4.24 & 4.25).

In general, these observations refer to a difference in transcriptional landscape between WT and RNase H-depleted cells. This situation can be interpreted as accumulation of R-loops on forward strand, as a result of RNase H deletion, may have suppressed transcription and R-loop formation over reverse strand. Surprisingly, this suggests a role of RNase H in suppressing forward strand R-loops and maintaining sense transcription over sense genes. Excitingly, this may also suggest a role for RNase-H in regulation of sense antisense transcription as increase of R-loop formation frequency over one strand may decrease transcription on the other strand.

Over some genes, the increase in forward-strand R-loop signals, completely depleted reverse-strand signals (Figures 4.18 & 4.23). This suggests that R-loops can form on both strands of the same gene within specific limits, but their over-accumulation on one strand may inhibit transcription and R-loop formation on the other strand. In general, this supports an anticorrelation between forward- and reverse-strand R-loop formation, and probably, anticorrelation between sense and antisense transcription.

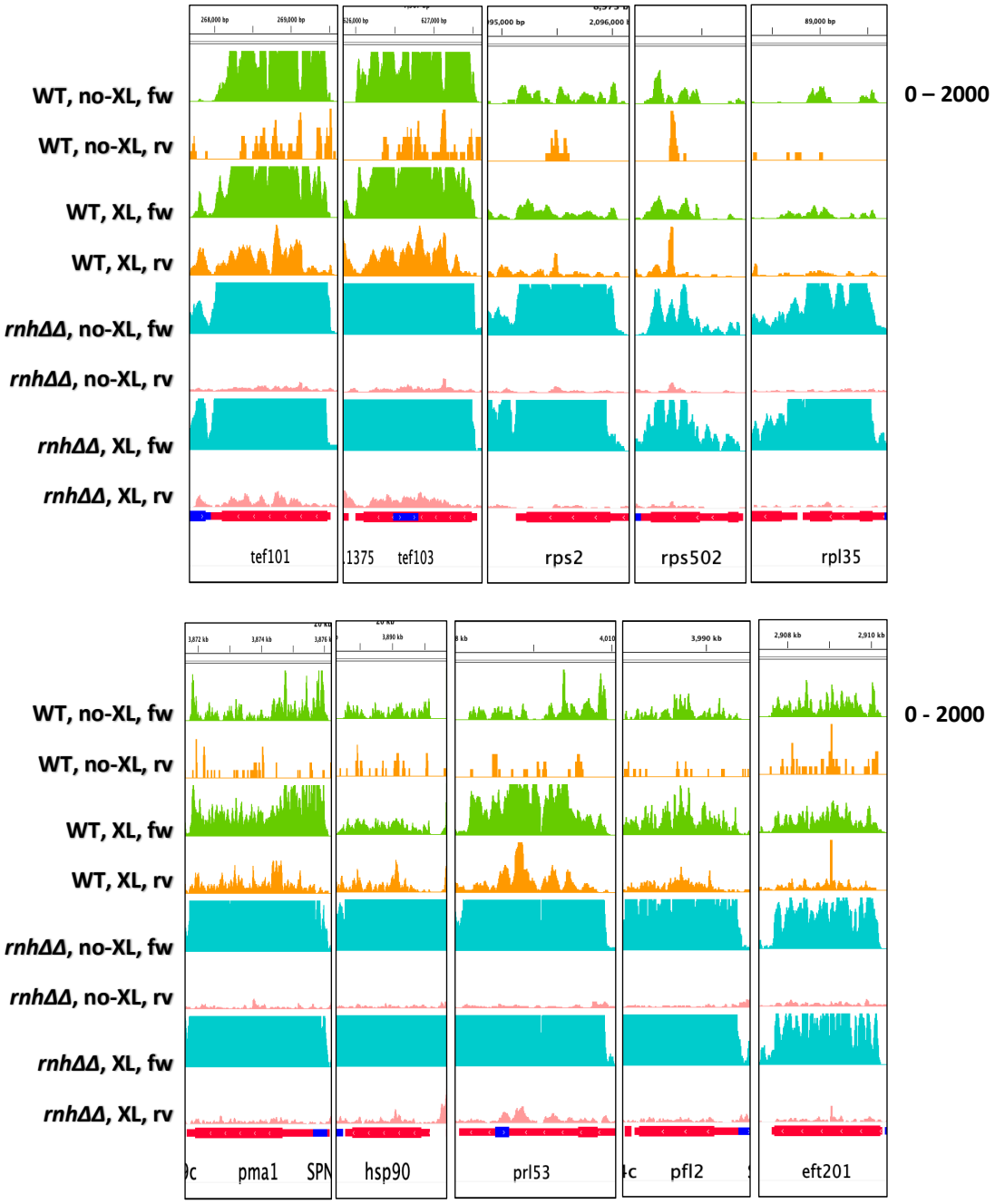



**Figure 4.23. RNase H deletion increased forward-strand while decreased reverse-strand R-loop signals.** Snap shots of genome browser, showing R-loop profiles over template and non-template strand of antisense genes.

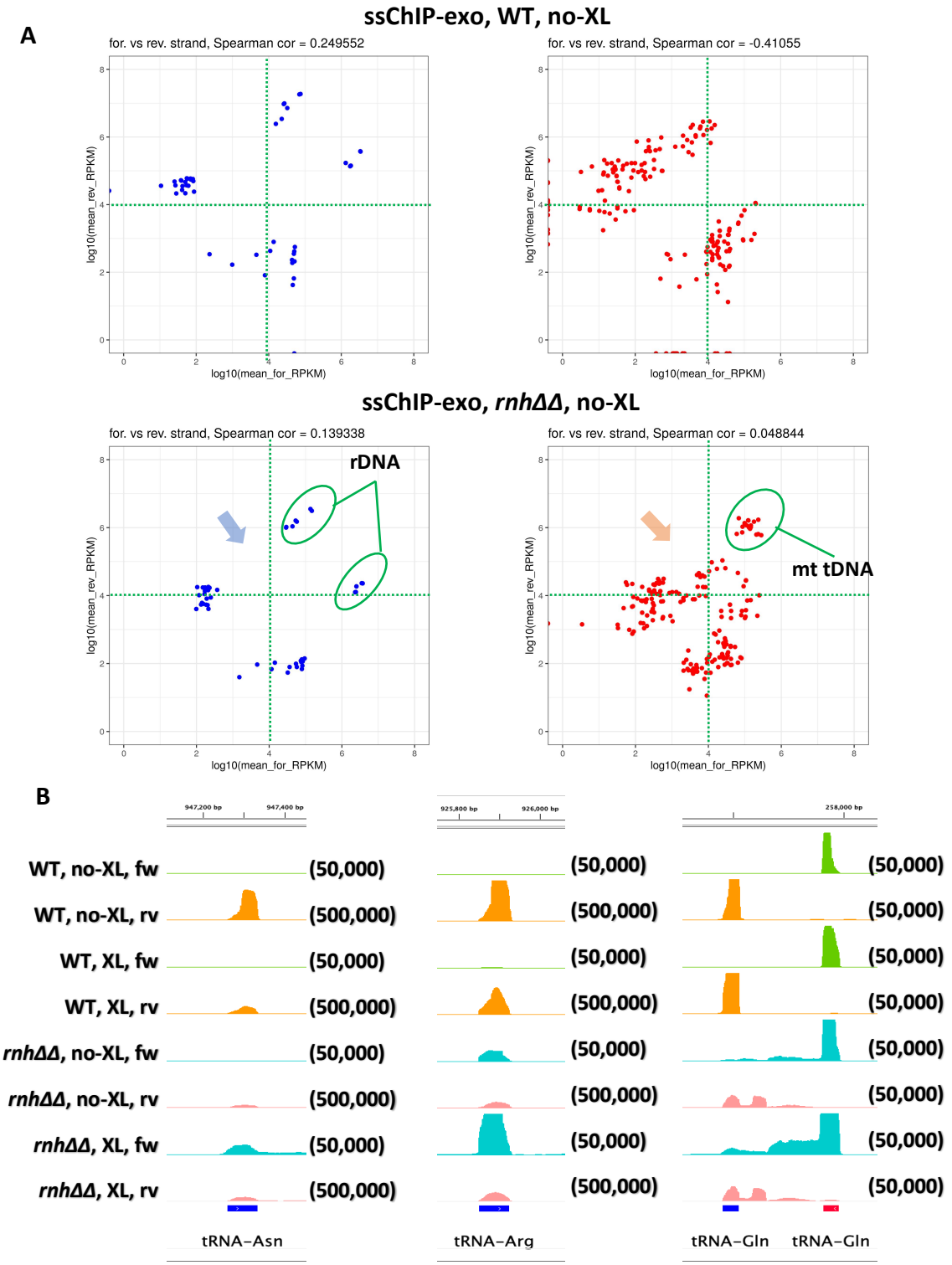



**Figure 4.24. RNase H deletion increased forward-strand while decreased reverse-strand R-loop signals over rDNA and tDNA repeats. (A)** spearman correlation analysis for forward- and reverse-strand R-loops signals over tRNA and rRNA genes. **(B)** snap shots of genome browser showing representative examples from tRNA genes. This effect for RNase H deletion was better seen over sense genes showing effect on both sides, while antisense genes show an effect on one side only. Compare similar conditions together. Arrows refer to direction of change compared to WT cells. Notice scaling.

### 4.3.16 Impact of RNase H deletion is reproducible for crosslinked cells

Remarkably, crosslinked cells exhibited the above effect for RNase H deletion on forward- and reverse-strand R-loop signals. The most remarkable effect for crosslinking was that it enhanced non-template strand signals over some regions including tRNA and rRNA genes (Figure 4.13). This only increased the correlation between forward and reverse strand signals, but didn't affect the consistency of the results or the final conclusion about R-loop formation over these regions (compare correlations in Figures 4.24 & 4.25). In agreement with a previous study which reported that crosslinking doesn't affect consistency of results of chromatin-based DRIP methods **(Halasz *et al.*, 2017)**, I found that crosslinking doesn't affect any of the results obtained, neither for dsChIP-exo nor for ssChIP-exo. Main results and conclusions are depicted and summarized in Figure 4.26 based on observations form ssChIP-exo.

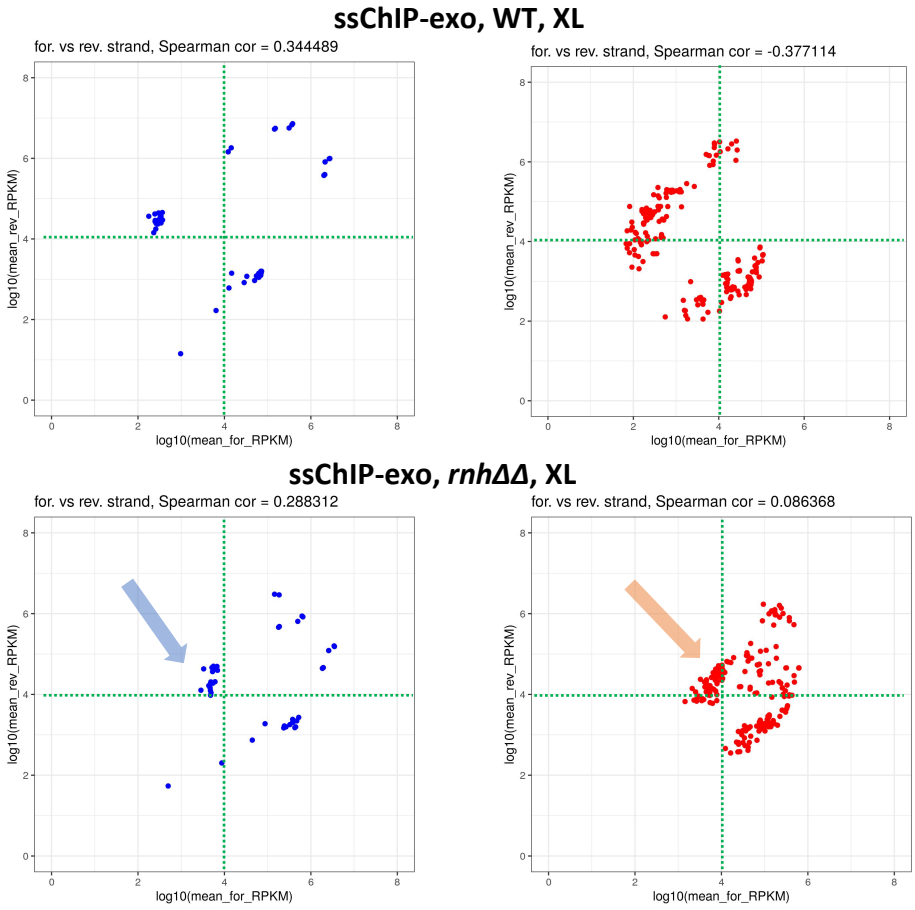



**Figure 4.25. Effect of RNase H deletion on R-loop signals of both DNA strands is reproduced for crosslinked samples.** Crosslinking increases the correlation between forward and reverse strand signals as it enhances non-template strand R-loops signals. Compare to correlations in Figure 4.24.

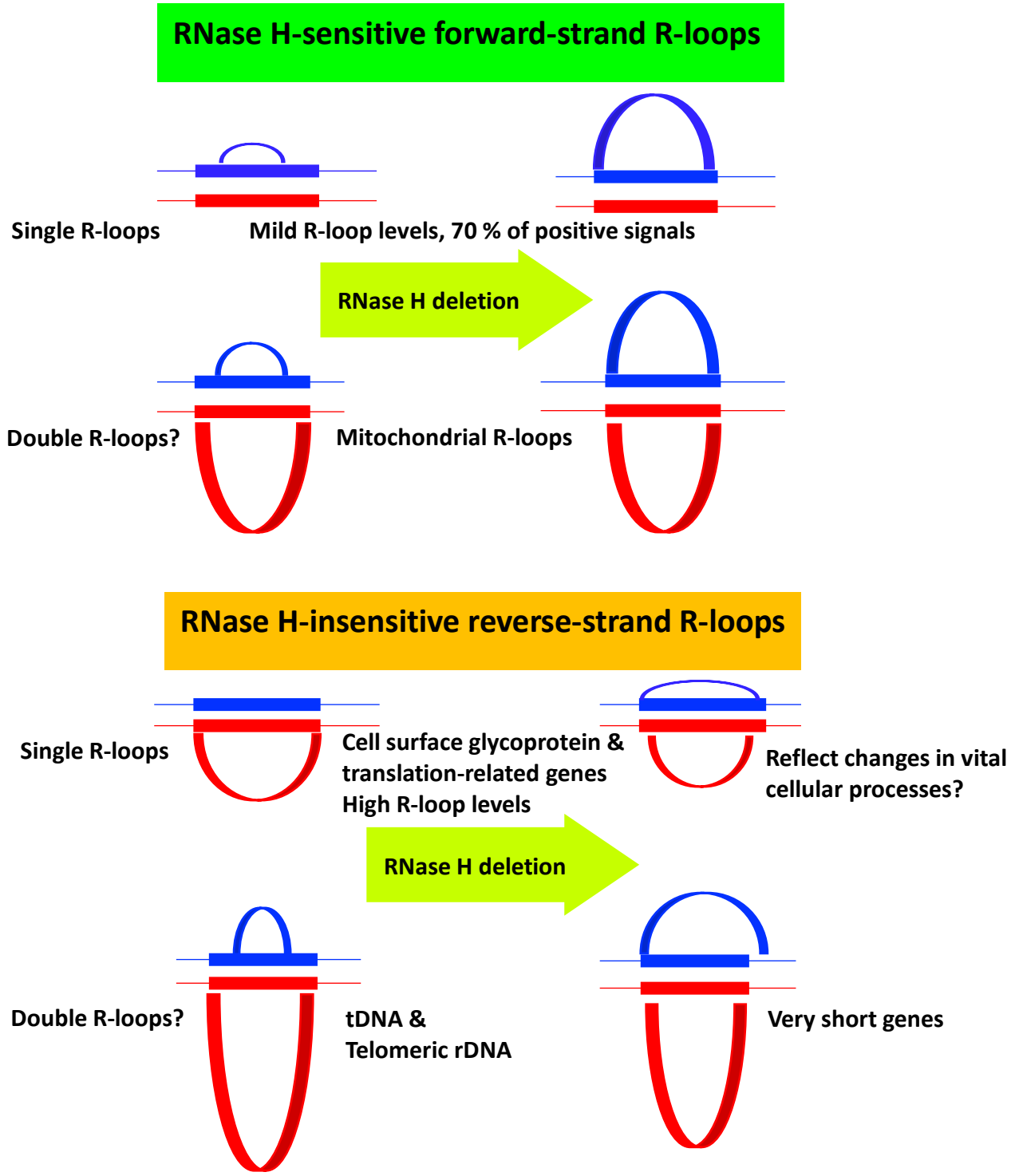



**Figure 4.26. Schematic representation for main R-loop categories identified based on sensitivity to RNase H and formation direction**. Depiction is based mainly on ssChIP-exo results for WT and *rnh1Δ rnh201Δ* cells. R-loop signals over 60 to 70% of positive loci are sensitive to RNase H. All regions, except tDNA, mitochondria, and telomeric rDNA, show single and co-transcriptionally directional signals. Signals of some stable R-loops over translation and ribosome biogenesis genes are decreased with RNase H depletion, implying change in transcriptional landscape. RNase H deletion increases signals over forward strand while decreases signal over reverse one. RNase H-insensitive signals map to reverse strand. For mitochondria, RNase H deletion severely affects forward-strand but not reverse-strand signal.

### 4.3.17 Preparation of permanently stable and three-stranded R-loops

Here I describe a method for preparing permanently stable and three-stranded artificial R-loops without the need for non-physiological denturing agents, DNA-modifying chemicals or non-canonical RNA bases. The use of spike-in standards for normalization has been previously suggested for genome-wide studies to replace normalizations based on total read count which may lead to bias in signal detection **(K. Chen *et al.*, 2015)**. Hence, the use of spike-ins seems to be more critical for R-loop mapping. It has been recently shown that the use of spike-ins allowed for more precise quantitation for R-loops levels under different biological conditions **(Crossley *et al.*, 2020)**. In addition to normalization, a mimic spike-in could be used for R-loop mapping to confirm directionality of signals and verify results from different methods. Spike-ins have been previously used with some R-loop mapping methods such as S1-DRIP **(Wahba *et al.*, 2016)** and qDRIP **(Crossley *et al.*, 2020)**. However, these spike-ins consisted only of RNA-DNA hybrids, not full R-loop structures which may not be the optimal control for all applictions. On the other hand, synthesis of full artificial R-loops can be also challenging as it requires the use of either formamide or high temperature for denaturing, and long incubation times for annealing, which may affect the integrity of these structures (Chapter 1, Section 1.4.2). It's also hard to confirm the persistence of these structures after diluting formamide by adding components to experimental solutions.

For artificial R-loop preparation, I tried to use near physiological conditions (without modifying or denaturing agents) for annealing two 500-nt ssDNA fragments (Fw & Rv) of random complementary sequence (Table 2.8) and *in vitro* transcribed 120-nt RNA (Figure 4.27A) that would form a median 120-bp hybrid with Rv DNA strand leaving flanking dsDNA of 190 bp at each side. First, I tried to prepare R-loops by direct annealing of the three strands after heat denaturation, which produced a band at 500 bp mark, the expected size for dsDNA (Figure 4.27B). This band was cleaved by restriction digestion suggesting that R-loop peparation isn't successful (Figure 4.27B). Alternatively, I annealed both RNA and DNA strands and efficienetly prepared an RNA-DNA hybrid before adding the second DNA strand. I found that the second DNA strand displaces the RNA of the hybrid to form dsDNA which was resistant to RNase H (Figure 4.27C) but sensitive to restriction digestion (Figure 4.27D).

These results don't match the concept that RNA-DNA hybrids, in general, are more stable than dsDNA and suggest that R-loop formation doesn't seem to be a direct process, but

requires special conditions. Also, additional factors may be required to stabilize R-loops *in vivo,* especially at high-transcription rate genes with low GC content. My conclusion is in line with previous studies which concluded that R-loop formation process is a complex reaction whose rate is a function of many factors, of which are reaction temperature, RNA concentration and ionic strength **(Birnstiel *et al.*, 1972; Thomas *et al.*, 1976)**. These two studies found that the stability of R-loops depend on features of RNA strand of the hybrid, these features include sequence composition (base content), molecular weight (RNA length) and secondary structures. For example, high molecular weight (long) RNA-DNA hybrids are more stable than low molecular weight (short) RNA-DNA hybrids, and hybrids with RNA of high GC content are stable compared to those with RNA of low GC content. This explains why rRNA repeats with high GC content were frequently used for R-loop preparations **(Birnstiel *et al.*, 1972; Thomas *et al.*, 1976)**. Consistenlty, it has been suggested that RNA of more than 100 nt is required for efficient RNA-DNA hybrid fromation. This length was found to be required to stabilize R-loop structure against RNA strand displacement by DNA through branch migration **(Landgraf *et al.*, 1995b)**. It's noteworthy that I used RNA with 120 nt but with only 50% GC content. Perhaps increasing RNA length, GC content or ratio of purine bases may have increased the efficienecy of hybrid formation and enhanced its stability. It has been adequately shown that increasing purine content in RNA, regardless of its length, increases stability of hybrids compared to dsDNA as a consequence of change in conformation toward A-form of dsRNA (Chapter 1, Section 1.4).

Importantly, my results don't completely contradict with previous publications which suggested that RNA-DNA hybrids are thermodynamically more stable than dsDNA as they studied long and GC rich R-loops in presence of Formamide **(Birnstiel *et al.*, 1972; Thomas *et al.*, 1976)**, a buffer that destabilizes dsDNA interactions **(Blake & Delcourt, 1996; Fuchs *et al.*, 2010; Sadhu *et al.*, 1984)** therby increasing the efficiency of hybridization and thermodynamic stability of RNA-DNA hybrids. It has been shown that removing formamide decreases the stability of RNA-DNA hybrids, which leads to RNA strand displacement by DNA and restoring dsDNA **(Birnstiel *et al.*, 1972; Kaback *et al.*, 1979; Landgraf *et al.*, 1995b; Landgraf *et al.*, 1996; Thomas *et al.*, 1976)**.

Notably, an equilibrium state may exist between RNA and non-template DNA of R-loops that there is a mutual exchange between both strands where the RNA strand can invade the duplex and nucleate hybrid formation while the DNA strand can displace the RNA throught

branch migration process. Importantly, the exchange process between ssDNA and RNA was found to be dependent on $Mg^{+2}$ which when removed, accelerated the RNA displacement process rate by 52-fold leading to spontaneous restoration of dsDNA **(Landgraf *et al.*, 1995b; Landgraf *et al.*, 1996)**. The fact that I used a magnesium-free annealing buffer means that once formed, the hybrids will almost get destabilized to restore dsDNA. This supports my observations from R-loop preparation trials above.

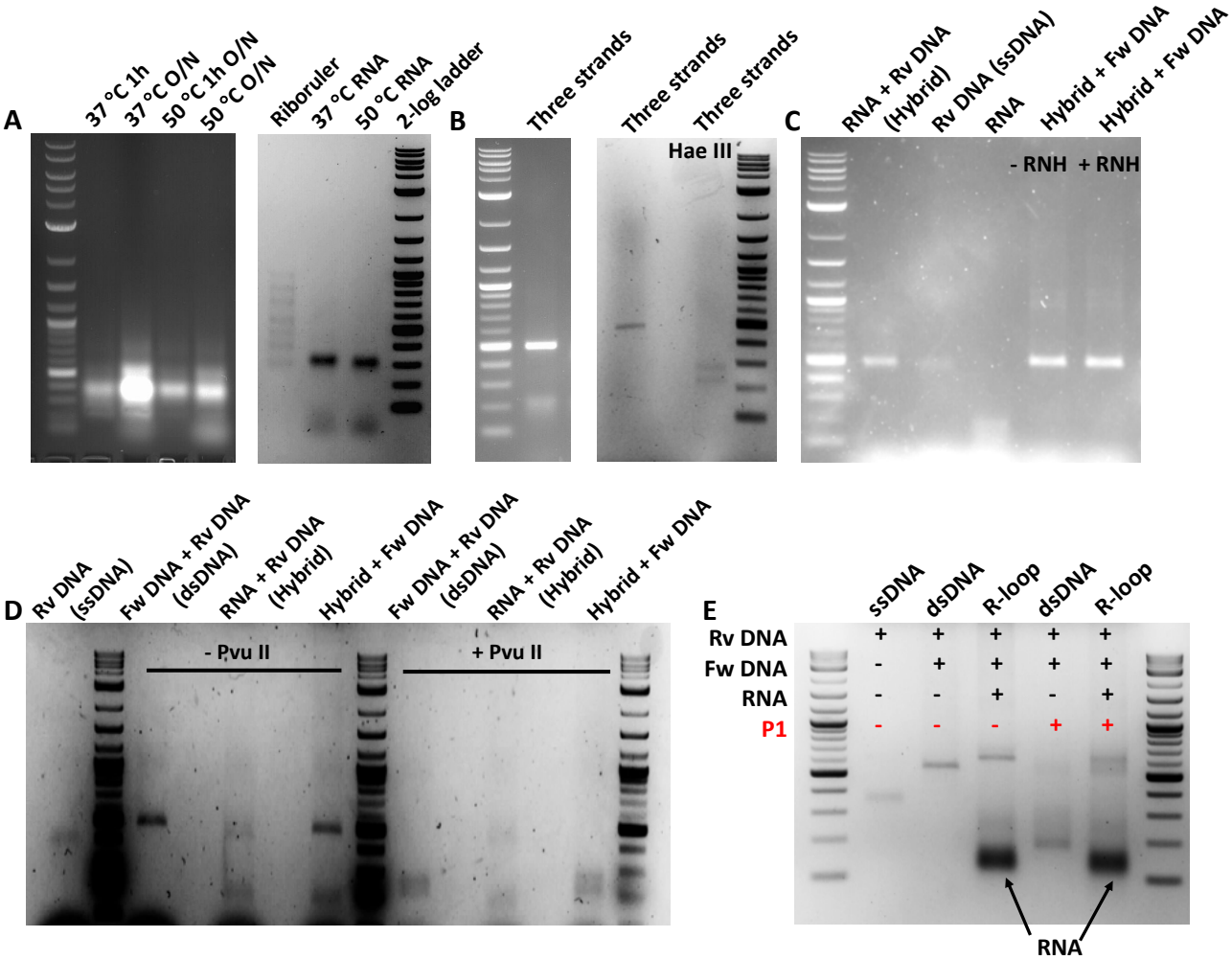



Figure 4.27. **Preparation for stable three-stranded artificial R-loops.** 3% agarose gel analysis for annealing components. **(A)** RNA *in vitro* transcribed using different conditions (left) and the same denatured RNA (right). **(B)** annealing of three strands (Fw & Rv DNA and RNA) produces a band that's sensitive to restriction digestion. **(C)** adding Fw DNA strand to annealed hybrid produces a band that's not sensitive to RNase H. **(D)** adding Fw DNA strand to annealed hybrid produces a band that's sensitive to restriction digestion. **(E)** successful preparation for R-loops using bubble dsDNA and RNA. Annealing product (R-loop) runs higher than dsDNA. P1 nuclease treatment cleaves bubble dsDNA into small fragments, different from R-loop.

For these concerns, I planned to use a different strategy for artificial R-loop preparation. This strategy depends on the use of two DNA strands with noncomplementary nucleotide sequence to form a bubble dsDNA after their annealing. This bubble dsDNA would allow

smooth hybridization of RNA to one of the DNA strands without any need for high temperature or special chemicals (Figure S4). This kind of R-loop would be helpful to confirm directionality of methods and test the source of non-template strand R-loop signals detected by methods like current dsChIP-exo and previous S1-DRIP. For this, I used two ssDNA fragments that would generate dsDNA with a bubble after annealing. The forward strand belongs to a random sequence while the reverse belongs to mammalian B-actin gene terminator, previously used and known for its R-loop formation potential **(Tan-Wong *et al.*, 2019)**. It has a high GC content and, importantly, low sequence homology with pombe genome. I used this B-actin DNA strand to generate complementary RNA using *in vitro* transcription. After confirming the integrity and right size of RNA, I annealed the three components directly in a very simple and short-time incubation. To verify the final product, I treated the annealed structure with P1 which digested only the bubble dsDNA but not the full R-loop (Figure 4.27E).

Beside prospective use for normalization and mimicking, this R-loop can be used for generation of double R-loops or multiple non-canonical structures. It can be used to test effect of different nucleases and to verify the ssDNA transcription model *in vivo* and *in vitro.* The fact that both DNA strands, at the hybrid region, have different sequences would be very helpful for directional mapping, accurate strand specific PCRs and multiple tests. Multiple spike-ins may be required for each step of the DRIP to test the effect of each step on R-loops (sonication, RNase H treatment, *etc*...).

## 4.4 Discussion

Here, I discuss reasons for choosing the current experimental procedures and conditions, and the suitability of different methods for directional R-loop mapping. Moreover, I discuss RNase-H control related issues and show the effect of different experimental conditions on R-loop profiling in light of my results. Furthermore, I discuss factors driving my investigation for a role of R-loops in dsRNA formation, and show my results in relation to recent knowledge. I also discuss limitations and challenges for suggested role of R-loops. Finally, I discuss implications of my findings regarding differential response of R-loops over both DNA strands toward RNase H deletion. This was mainly enabled by the enhanced directionality of the new ssChIP-exo method.

### 4.4.1 R-loops, antisense transcription and potential dsRNA formation

A growing body of evidence suggests formation of dsRNA through sense antisense transcription in yeast and mammalian cells. Intriguingly, R-loops have been suggested to contribute to dsRNA formation and, also, to regulation of gene expression through the RNAi-dependent heterochromatin machinery **(Skourti-Stathaki *et al.*, 2014)**. Generation of dsRNA by R-loops, if possible, intuitively, suggests the abundance of two complementary RNA strands, at least, one of them gets engaged in RNA-DNA hybrid formation. These two RNA strands may be produced either by sense antisense transcription from the same locus (*in cis*) or from two loci of homologous sequences (*in trans*). R-loop-dependent *in cis* dsRNA generation is the focus of my work in line with the fact that *in cis* R-loops represent the most dominant type of R-loops in different organisms.

Theoretically, different pathways can be envisaged for an R-loop to form dsRNA. The first pathway may involve the RNAi machinery which is conserved in fission yeast and higher eukaryotes. dsRNA generation by this machinery relies on RNA signal amplification by the RNA-dependent RNA polymerase (RDRP). Practically, this suggestion doesn't seem to be adequately supported by scientific findings as RDRP hasn't been shown to localize at euchromatic regions or at loci ectopic to heterochromatic domains. Interestingly, Dcr1, a member of the RNAi machinery has been shown to have a role in Pol II transcription termination and release at a subset of Pol II and Pol III gene terminators in fission yeast. Importantly, these regions are correlated with Pol II pausing and represent replication fork-transcription collision sites. However, none of the RNAi machinery was found to participate in this role along with Dcr1, confirming that this role is specific for Dcr1 **(Castel *et al.*, 2014)** independent of the RNAi machinery. Intriguingly, defects in termination, as a result of Dcr1 deletion, led to accumulation of deleterious R-loops, probably, through increasing transcription replication collision. However, neither Dcr1 nor RDRP has been linked to a role in processing these R-loops **(Castel *et al.*, 2014)**. Mechanistically, amplification of RNA via RDRP requires the RNA to be single stranded, which excludes the need for R-loops, albeit for stabilizing the RNA.

It needn't say that the genome of different organisms is pervaded by different forms of antisense transcripts. Antisense long noncoding RNA (aslncRNA) have been found to form extensively over more than 70% of fission yeast genome and over the genome of mammalian

cells **(Tan-Wong *et al.*, 2019; Wery, Gautier, Descrimes, Yoda, Migeot, *et al.*, 2018; Wery, Gautier, Descrimes, Yoda, Vennin-Rendos, *et al.*, 2018)**. It's really tempting and, also, safe to think that with this huge bulk of antisense transcripts, the antisense-RNA-producing loci may be an R-loop forming region as well. The presence of two complementary RNA strands may elicit possible interactions between RNA of an R-loop and asRNA and, eventually, dsRNA formation. Indeed, R-loops have been found to be associated with antisense transcription in budding yeast *S. cerevisiae* **(Chan *et al.*, 2014)** and in mammalian cells **(Skourti-Stathaki *et al.*, 2014; Tan-Wong *et al.*, 2019)**.

Importantly, dsRNA pairing may be favoured by higher thermodynamic stability of RNA-RNA compared to all corresponding nucleic acid pairs (RNA-DNA or DNA-DNA) of the same sequence **(Hall & McLaughlin, 1991; Hung *et al.*, 1994; Kankia & Marky, 1999; Lesnik & Freier, 1995; Ratmeyer *et al.*, 1994; Roberts & Crothers, 1992; Sugimoto *et al.*, 1995)**. Mechanistically, at R-loops, RNA-DNA hybridization may represent a barrier against dsRNA formation at physiological conditions, especially at biological temperature which isn't enough to break RNA-DNA associations, i.e., RNA has to be separated first from DNA before pairing with the asRNA for dsRNA formation. This doesn't seem to be a problem knowing that it has been shown that the RNA strand can be displaced by DNA through a branch migration process **(Landgraf *et al.*, 1995a, 1995b; Landgraf *et al.*, 1996)**. Also, cellular enzymes such as RNA-DNA helicases can facilitate unwinding of RNA-DNA hybrids and separation of RNA. This is supported by the fact that loss of SETX or DHX9 helicases was found to accumulate R-loops and perturb transcription termination **(Cristini *et al.*, 2018; Skourti-Stathaki *et al.*, 2011)**. Interestingly, it has been shown that *in vitro* bidirectional transcription at an R-Loop forming locus, in a plasmid template, can lead to dsRNA formation in absence of RNA-DNA helicases **(Hartono *et al.*, 2018)**. This final finding shows the feasibility of dsRNA formation by R-loops. However, it's not known whether the presence of chromatin would affect these resolution annealing processes.

Although the capability of R-loops to form dsRNA hasn't been widely investigated, the potential of asRNAs to form dsRNAs has been recently uncovered. Intriguingly, an elegant study has shown that asXUTs, a subset of antisense transcripts sensitive to Xrn2/Exo2 subunit of the cellular exosome, have the potential to anneal with the sense RNA and form dsRNA. Using the RNAi-lacking *S. cerevisiae*, but with artificially reconstituted functional RNAi machinery (exogenous expression of dicer and ago1), this study reported accumulation of small

RNAs with the majority mapping to asXUTs which are detected after deletion of xrn2 **(Wery *et al.*, 2016)**. Recently, a subset of antisense transcripts has been found to be sensitive to Dcr1, implying the formation of dsRNA through annealing with paired sense transcripts. However, knowledge about this role is still limited to specific genomic regions **(Atkinson *et al.*, 2018)**. Intriguingly, even in WT budding yeast lacking the dsRNA-recognizing Dcr1 and the RNAi machinery, aslncRNA transcripts have been implicated in dsRNA formation through the nonsense mediated decay (NMD) process. This involves formation of dsRNA through annealing of paired sense antisense transcripts which contain a 3' single stranded overhang protruding from the ncRNA. This overhang is required for recognition and unwinding by two specific helicases **(Wery *et al.*, 2016)**. This mechanism is thought to be an alternative for the RNAi in fission yeast and higher eukaryotes. Importantly, these findings show that dsRNA formation isn't limited to RNAi machinery.

Here, I proposed two models explaining possible R-loop-mediated dsRNA formation mechanisms. These models suggest that antisense RNA transcribed over the same R-loop forming locus, may anneal to the sense RNA of RNA-DNA hybrid and form dsRNA. However, these models differ in the form the antisense RNA; it may stay intact with the DNA and form an antisense RNA-DNA hybrid in a structure that can be called a double R-loop; or alternatively, the antisense RNA may exist in a free form after getting promptly and co-transcriptionally separated from the DNA. The first model of double R-loop formation gains support from the fact that RNA-DNA hybrids form *in vivo* over both ssDNA ends of double strand break (DSB) **(Ohle *et al.*, 2016)** and over transfected single-stranded DNA gaps and overhangs in fission yeast (unpublished data, Fischer lab). These results suggest that, likewise, RNA Polymerase can transcribe free ssDNA of R-loops to form antisense RNA-DNA hybrid.

It is noteworthy that this work was also driven by results from previous ChIP-exo data that supported genome-wide double R-loop formation. However, based on previous findings that different methods can produce different and inconsistent results, I decided to study the reproducibility of these results using different methods, cell types, experimental conditions and appropriate controls.

### 4.4.2 ssChIP-exo and SMART-DRIPc; new methods to overcome limitaions of ssDRIP and DRIPc

Unfortunately, the lack of resolution and directionality of methods used for R-loop mapping didn't facilitate understanding the associations between R-loops and antisense transcription which are still elusive. The precise localization and direction of R-loop formation, in general, and in relation to the corresponding asRNAs, in particular, aren't clear. For example, it's not clear whether, and to what extent, the non-template strand can form a hybrid. Also, the correlation between the template and non-template strand hybrids and the correlation between R-loop formation and antisense transcription aren't clear.

One of the major aims of this work and, also, the first step toward examining the different proposed models, was to deliver a specific, high resolution and, importantly, directional R-loop mapping. Therefore, I used three different methods to map R-loops in *S. pombe* as a model organism for the current study. For the sake of avoiding inconsistency among the three methods, I adopted the same DRIP approach, the most widely used for R-loop mapping, for all the methods. The earliest edition of this approach depended on the use of S9.6 antibody for the immunoprecipitation of R-loops from fragmented nucleic acids, ending by either microarray analysis or next generation sequencing for the purified DNA eluate. The classic DNA-based DRIP, which is based on the sequencing of dsDNA template, has critical limitations standing against precision analysis, which are lack of resolution and absence of directionality. This means that the exact R-loop forming sequence and the direction of R-loop formation can't be precisely defined. For this reason, different modifications and versions have been introduced by different labs to improve different aspects of the DRIP method, especially resolution and directionality. The best known of these versions are: 1) RDIP method which involved the use of sonication, RNase I treatment and DNA-extension using nicked RNA as primers for directional mapping of R-loops in mammalian cells **(Nadel *et al*., 2015)**: 2) DRIPc which introduced sequencing of the RNA instead of the DNA strand of the hybrid for directional profiling **(Chedin *et al*., 2016)**: 3) S1-DRIP which used S1 nuclease to digest displaced DNA strand and map only the hybrid-forming DNA strand for budding yeast **(Wahba et al., 2016)**: 4) ssDRIP to map R-loops in *Arabidopsis* through sequencing the DNA strand involved in hybrid formation, using a single stranded DNA ligation as a ssDNA-capturing and directional approach **(Xu *et al*., 2017)**: and 5) BisDRIP, a method that used bisulphite conversion in

combination with S9.6 immunoprecipitation to present high resolution and directional mapping.

A distinguished variant of the DRIP method is the ChIP-DRIP or DRIP-ChIP which uses sonicated chromatin instead of total nucleic acids extract, for immunoprecipitation by the S9.6 antibody. For this approach, crosslinking is often done before shearing and immunoprecipitation. This method has the same limitations of the classic DRIP. In a trial to increase resolution and enhance directionality of ChIP-DRIP, the ChIP-exo method was used for mapping R-loops in fission yeast, exploiting the power of lambda exonuclease that digests 5' ends of dsDNA **(Ohle *et al.*, 2016)**. Although this method enhanced resolution, the directionality wasn't verified. Notably, during my current work, I investigated the directionality of this method against other methods.

Independent from S9.6-based R-loop mapping methods, another category of methods for R-loop mapping was developed to depend on the use of *in vivo*-expressed and catalytically dead epitope-tagged RNase H for capturing R-loops. This has been achieved through immunoprecipitating RNase H using anti-epitope antibody for indirect mapping of bound R-loops. The best known for these are the R-ChIP and RR-ChIP. R-ChIP ends by sequencing the ssDNA while RR-ChIP sequences the RNA moiety. The biggest advantage is that they map quickly degraded and native dynamic hybrids of short life time **(L. Chen *et al.*, 2017; Tan-Wong *et al.*, 2019)**.

Considering the caveats of the different methods, S1-DRIP couldn't deliver directional mapping, instead, the produced signals resembled canonical dsDNA signals **(Wahba et al., 2016)**. Apparently, bisDRIP is the most complicated requiring a lot of replicates, controls and complicated analysis. Additionally, bisDRIP has the potential to show R-loops signals for single-stranded DNA at promoters and other regions that don't form real R-loops **(Dumelie & Jaffrey, 2017)**. Besides, its resolution is dependent on the presence of cytosines over the free displaced ssDNA. The biggest caveat for bisDRIP is that it depends on the presence of ssDNA for detetction of signal through bisulphite conversion which means this method won't be able to detect double R-loops as there will be no free ssDNA, so in this case, only S9.6 signal will be detected, but not bisDRIP footprint. R-ChIP and RR-ChIP were found to detect fewer R-loops compared to the DRIP-like methods **(L. Chen *et al.*, 2017; Tan-Wong *et al.*, 2019)**. The reason is that a subset of these R-loops may not be targeted by RNase H, but rather processed

by other ribonucleases that are abundant in the nucleus **(Vanoosthuyse, 2018)**. In addition, it may not be able to capture short hybrids compared to DRIP methodology which is based on S9.6 that can bind short hybrids as to 6 base pairs **(Phillips *et al.*, 2013)**. This may require enriching of short fragments or use of specific library preparation, though. Finally, R-ChIP and RR-ChIP may have some drawbacks as they may irreversibly bind to their target R-loops and interrupt their resolution, which may block transcription, thereby inducing undesirable artificial effects **(Tan-Wong *et al.*, 2019)**.

Collectively, I found both ssDRIP and DRIPc methods to be the best candidates to be used. ssDRIP and DRIPc presented substantial improvement over early methods as they delivered directional mapping and stranded signals. Importantly, using ssDNA- and RNA-based methods would be complementary in the sense that their signals should overlap and results should be confirmative. From the directionality perspective, the DNA signal captured by ssDNA-based should complement (form opposite to) the RNA signal of RNA-based method. Yet, the resolution of ssDRIP is still dependent on fragment size of DNA. In addition, single-stranded adaptor ligation used for ssDRIP may represent a limitation against sensitivity and lead to underrepresentation of some signals. Similarly, free unfragmented RNA may affect the resolution of DRIPc. Moreover, working with RNA may limit the sensitivity of DRIPc and lead to underrepresentation of R-loops signals at some regions. However, these drawbacks can be eliminated using dedicated experimental procedures (next).

To overcome the low-resolution problem and eliminate the dependence of resolution on DNA fragment size, I decided to use the ChIP-exo methodology, a modification of the ChIP technique. Beside sonication which produces small fragment size, this method uses lambda exonuclease that digests the 5’ end of DNA till reaches the crosslinking or DNA binding site, which is the end of RNA-DNA hybrids in case of R-loops. This has been found to greatly reduce the background and deliver a near-single nucleotide resolution mapping. Another exonuclease step is incorporated to digest unbound single stranded DNA. This step minimizes the background signal as well.

First of all, I used the original ChIP-exo method which depends on double stranded adaptor ligation as a dsDNA-capturing method, so I called this method as dsChIP-exo. Mapping R-loops using this method was important to compare general mapping results to those previously produced by a similar method **(Ohle *et al.*, 2016)**. In parallel with dsChIP-exo, and

similar to ssDRIP principle, I used ChIP-exo method in combination with a ssDNA-capturing method to deliver a high resolution and directional mapping. Instead of ssDNA adaptor ligation, I adopted dA-tailing for higher sensitivity. I called this method ssChIP-exo (the same as ChIP-exo-TT).

To confirm and support directionality of ssChIP-exo, I followed DRIPc footsteps for an RNA-based mapping method. A problem with DRIPc, the best known RNA-based method, that has been encountered in fission yeast **(Hartono *et al.*, 2018)**, but not in mammalian cells **(Sanz & Chedin, 2019; Sanz *et al.*, 2016)**, is dsRNA contamination, a consequence for affinity of S9.6 for dsRNA **(Hartono *et al.*, 2018)**. In order to enhance resolution and decrease dsRNA contamination, I treated the RNA with two dedicated nucleases, RNase III and P1 nuclease, to digest dsRNA and ssRNA, respectively. Finally, to enhance sensitivity and at the same time directionality, I used the SMART technology for the library preparation step. I called the final method as SMART-DRIPc. Different form external/classic DRIPc which depends on reverse transcription and cDNA synthesis using random primers, SMART-DRIPc cDNA synthesis step is performed using a primer with a linker to avoid non-specific priming and chimeric spurious reverse transcription **(Haist *et al.*, 2015; Z. Peng *et al.*, 2015)**.

Overall, I mapped R-loops using three different methods, two of them are DNA based and directed to sequence the DNA while the third is RNA-based and directed to sequence the RNA moiety. For the DNA-based, I used the original ChIP-exo technique (here called dsChIP-exo) as a high-resolution method to capture the dsDNA in the first method. In the second method, I used the ChIP-exo-TT (here called ssChIP-exo) as a ssDNA-based method to capture only the hybrid-forming DNA strand and enhance the directionality of the signals. As a complementary method for the second one, I used the DRIPc method in combination with the SMART technique (here called SMART-DRIPc) for sensitive capture of the RNA. For the SMART-DRIPc method, I included different enzymatic treatments to digest any dsRNA and free ssRNA contamination.

Remarkably, my three methods exhibited higher quality, compared to external ones, represented by lower background noise, higher resolution and enhanced specificity. My methods showed a great overlap among their signals, and compared to other external methods. I found that fixation may variably affect enrichment of signals over specific loci and also the total signal for WT and RNase H depleted cells, but importantly, fixation didn't affect the final

results or conclusions. Different from dsChIP-exo, ssChIP-exo and SMART-DRIPc showed unidirectional signals suggesting that dsChIP-exo isn't compatible for directional R-loop mapping. SMART-DRIPc didn't capture dsRNA or show double signals even without enzymatic treatments, suggesting that it's inherently specific and directional. Observations from ssChIP-exo and SMART-DRIPc regarding signal directionality don't support double R-loop formation. Furthermore, most of R-loops were sensitive to RNase H in WT which weakens R-loop-dependent dsRNA formation models in general. On the other hand, unidirectionally-formed RNase H-insensitive R-loops correlated with asRNA and dsRNA signals suggesting that second dsRNA formation model is more plausible. Finally, RNase H deletion increased signal over forward strand, but decreased them over reverse strand, suggesting an impact on transcription of sense genes. This effect was observed over sense-transcribed ribosome biogenesis and translation-related genes. These results suggest that RNase H may affect cellular transcriptomic and proteomic profiles independent from R-loops.

### 4.4.3 Crosslinking induces variations in R-loop signals, but doesn't affect final results

Remarkably, formaldehde fixation increased cumulative R-loop levels for WT cells. This was revealed by enrichment of total PCR signals using dsChIP-exo and ssChIP-exo methods for crosslinked compared to non-crosslinked WT cells (Figures 4.6 & 4.7). This finding is consistent with the fact that fixation may induce conformational changes and enhance dsDNA denaturation **(McGhee & Von Hippel, 1977)** which may induce R-loop formation through increasing potential for RNA-DNA hybridizatin. On the contrary formaldehyde fixation reduced R-loop levels for *rnh1Δ rnh201Δ* cells as PCR signal was weaker for crosslinked compared to non-crosslinked *rnh1Δ rnh201Δ* cells. Also, crosslinked *rnh1Δ rnh201Δ* cells showed lower enrichment for some mapped signals compared to non-crosslinked *rnh1Δ rnh201Δ* cells. This was also revealed by number of signal peaks detected for crosslinked and non-crosslinked *rnh1Δ rnh201Δ* cells by dsChIP-exo and ssChIP-exo methods. These observations can be explained by a previous finding that crosslinking weakens recognition of R-loops by S9.6 antibody **(Legros *et al.*, 2014)**.

The fact that WT signals strongly increase with crosslinking clearly confirms that formaldehyde may induce R-loop formation, probably through destabilizing dsDNA interactions or decreasing strands re-association in a way that might be similar to formamide (Section 1.4.). On the contrary, results for *rnh1Δ rnh201Δ* cells show that crosslinking affects

the signal negatively. My evaluation is that WT cells have low R-loop levels because R-loops are constantly degraded by RNase H. Formaldehyde fixation may induce R-loop formation at sequences with potential for R-loop formation, i.e., the same sequences which form R-loops in RNase H-deleted cells (R-loops which increase by RNase H deletion). On the other side, RNase H-deleted cells are already saturated with R-loops, i.e., R-loops form over all possible sequences (which have the potential for R-loop formation) so that fixation can't induce more R-loops. This means, the induction effect for fixation won't be seen for RNase H-deleted cells, but only the blocking effect or the negative effect on recognition by antibody, which is the reason for decreased signal in these cells. Apparently, crosslinking seems to have a double-way effect. It enhanced signals over some regions while weakened signals over others. The reason for this differential effect isn't clear.

Altogether, formaldehyde fixation has two different effects; it may induce hybrid formation over some sequences; and at the same time, it may inhibit affinity of S9.6 to RNA-DNA hybrids even in the same sample. The induction effect seems to be more obvious for WT cells which have low R-loops levels. On the other hand, RNase H-depleted cells have high (saturated) R-loops levels that formaldehyde can hardly induce more hybrid formation. That's why inhibition effect will be more observed for these cells.

Notably, crosslinking enhanced R-loops signals of non-template strand over rRNA and tRNA genes for both WT and *rnh1Δ rnh201Δ* cells compared to the non-crosslinked cells, but didn't affect the template strand signal strength. This slightly increased the correlation between forward- and reverse-strand R-loops over these genes (Figures 4.24 & 4.25). My explanation is that crosslinking increases the chance of capturing the free non-template DNA strand through fixing it to other proteins that may be also binding to the RNA-DNA hybrid. These proteins may get crosslinked to the whole R-loop, especially short R-loops, and increase the chance of capturing the free DNA strand to be processed for the library preparation.

Overall, crosslinking didn't affect the final results or conclusions regarding specificity, resolution, or directionality of signals, and, importantly, regarding correlation studied for forward- and reverse-strand R-loop signals (Figures 4.24 & 4.25). The main reason, in my opinion, is that the current mapping methods depend on the use of chromatin preparation as an input which may have limited the effect of formaldehyde. Importantly, a chromatin-based method would capture R-loops in more native chromatin context, maintaining the intactness of

R-loops and preserving the R-loop footprints of chromatin, i.e., decreasing strand breathing, shifting, denaturing and reannealing events during lysis and sonication steps. This is also the reason why RNase H-based methods, such as R-ChIP and RR-ChIP, were developed. In addition, nucleic acids extraction may require phenol chloroform extraction and ethanol purification that may affect R-loops **(Crossley *et al.*, 2020)**.

### 4.4.4 Lower noise, higher resolution and stronger overlap in current R-loop profiles

To validate the results of this study, I confirmed the high quality and reproducibility of my mapping profiles through comparing the background noise, resolution, overlap and localization of signals captured by my methods to those of external methods for *S. pombe.* Surprisingly, I found the data of my DNA- and RNA-based methods to have better quality, compared to corresponding public data, represented by higher resolution of signals and lower noise to signal ratio (Figures 4.9, 4.10 & 4.12). The low background noise can be confirmed by the absence of PCR amplification in no-antibody controls as seen in agarose gel and bioanlayzer profiles. Also, minimal noise was detected in genome browser tracks compared to external methods. The high resolution can be clearly confirmed against other data by inspecting signal ends, low noise and bioanlayzer profiles. Clearly, bioanalyzer profiles showed average fragment size of ~300 bp and ~200 bp for dsChIP-exo and ssChIP-exo amplified DNA, respectively, which suggests that ssChIP-exo may have higher resolution. All these observations confirm the specificity of the mapping which can be also confirmed by the sensitivity of PCR signals to RNase H treatment in all experiments and overlap of detected signals with R-loop formation hotspots of fission yeast.

Importantly, I detected strong spatial overlap among signals of my current methods and among signals of my current and external methods. Moreover, the overlap among signals of my methods was more pronounced than previously reported for other methods. Importantly, ~90% of the ssChIP-exo signal peaks overlapped with ~55% of the dsChIP-exo signal peaks for RNase H-deleted cells (Figure 4.9). Moreover, more than 75% of the RNIII/P1-treated SMART-DRIPc sample signal peaks overlapped with signal peaks from both SMART-DRIPc and external DRIPc (Figure 4.10). In particular, these observations suggest the reliability of ssChIP-exo and RNIII/P1-treated SMART-DRIPc samples profiles, since they captured signal peaks in common with other methods. Finally, ~50% of SMART-DRIPc signal peaks overlapped with 44% of dsChIP-exo (Figure 4.11).

Remarkably, this overlap represents the strongest consistency and overlap ever shown among different methods, knowing that previous studies mapping R-loops in mammalian cells reported an overlap of roughly 25% among methods **(L. Chen *et al.*, 2017; Tan-Wong *et al.*, 2019)**. To clear the sense of exaggeration in my conclusion, it's important to note that the percentage of overlap (around 50%) shown for dsChIP-exo signals represents the maximum overlap possible for this data. By other words, the dsChIP-exo can't show an overlap of more than half the number of its signals. The reason is that signals were predicted for forward and reverse strand separately, but, not predicted based on positive loci. Given the fact that there is a difference in directionality among methods (next section), a single locus may be presented by two signals (both sides of DNA) in case of dsChIP-exo, while the same locus will be presented by a single signal in case of ssChIP-exo and SMART-DRIPc. This strong overlap, along with other observations, support the reproducibility of my data and consistency among my methods. My observations also confirmed the fact that the more the difference among the experimental conditions and methods, the less the overlap and the similarity among their results.

Strikingly, signals detected by my methods mapped to known hot spots of R-loop formation in fission yeast. My data reinforced all previously known facts about R-loops strong accumulation over non-coding RNA Pol III and RNA Pol I genes, highly-transcribed protein-coding RNA Pol II genes, retrotransposon elements, heat shock proteins and mitochondrial repeats in agreement with reanalyzed data for fission yeast **(Bronner *et al.*, 2017; Castel *et al.*, 2014; Hartono *et al.*, 2018; Ohle *et al.*, 2016)**. Moreover, R-Loop mappings of my current study are consistent with other previous reports for *S. Cerevisiae* **(Chan *et al.*, 2014; El Hage *et al.*, 2014; Wahba *et al.*, 2016)**, *S. pombe* **(Hartono *et al.*, 2018; Legros *et al.*, 2014),** mammalian cells **(Crossley *et al.*, 2020; Ginno *et al.*, 2013; Ginno *et al.*, 2012; Nadel *et al.*, 2015; Sanz *et al.*, 2016)** and *Arabidopsis* **(Xu *et al.*, 2017)** about signal enrichment over the above regions and regarding R-loops accumulation over highly transcribed genes (Figure 4.13)

It may be argued that these methods, especially the SMART-DRIPc captures background contamination from highly expressed genes such as tRNAs and from total RNA in the cell. Two facts stand against this argument, the first is that S9.6 antibody has no affinity to dsDNA and that both the DNA-based methods (dsChIP-exo and ssChIP-exo) can't capture or sequence RNA. Detection of strong signals over transcriptional units and highly transcribed genes, particularly, support specificity of DNA-capturing dsChIP-exo and ssChIP-exo as they only

capture DNA which is supposed to have fixed copy number along the genome, regardless of transcription. Second, SMART-DRIPc R-loops profiles look different from RNA-seq profiles, which provide a clear-cut evidence that my methods are very specific and capture genuine R-loop signals. Finally, the use of P1 should confirm the digestion of free mRNA at least in the treated sample. For ssChIP-exo, there is no way for any DNA contamination doubt, as signals always belong only to one DNA strand. Beside striking enrichment of signals over known R-loop formation hot spots, other observations such overlap of signals among methods and sensitivity of signals toward RNase H confirm the specificity of the mapping.

### 4.4.5 ssChIP-exo and SMART-DRIPc, but not dsChIP-exo, show unidirectional signals

Metagenic analysis for dsChIP-exo data revealed similar distributions and strengths for forward- and reverse-strand metagenic signals, and overlap of their central metagenic peaks over forward and reverse strand (Figure 4.14). This suggested the wide formation of double R-loops, similar to a previously generated ChIP-exo data **(Ohle *et al.*, 2016)**. The same analysis for the ssChIP-exo and SMART-DRIPc data didn't reveal the same pattern casting doubt on the validity of the conclusions drawn from the dsChIP-exo (Figure 4.15). Careful inspection for individual R-loops signals of the different methods, showed that different form dsChIP-exo, both ssChIP-exo and SMART-DRIPc produced stranded signals forming mainly over the template DNA strand (Figure 4.13). These observations support formation of co-transcriptionally directional R-loops, which is consistent with knowledge in the field. In agreement with all directional R-loop mapping studies on different organisms, using different methods that showed very few regions with sense antisense R-loops **(L. Chen *et al.*, 2017; Hartono *et al.*, 2018; Sanz *et al.*, 2016; Tan-Wong *et al.*, 2019; Xu *et al.*, 2017)**, my results don't support the wide formation of double R-loops. Importantly, ssChIP-exo and SMART-DRIPc stand against dsChIP-exo results and refute the wide formation of double R-loops.

It's important to stress that all the versions of ChIP-exo method are strand-specific due to the use of different sequencing adaptors for the 5' and 3' ends of DNA. This enables the recognition of the different 5' ends and, more importantly, the separation of strand signals during bioinformatic analysis. Independently, the high resolution is conferred by the use of lambda exonuclease that cleaves the 5' end of dsDNA to the borders of protein-DNA interaction (in case of R-loops, this enzyme should cleave the DNA to the borders of the RNA-DNA hybrid). It's important to differentiate resolution and strand specificity from

directionality which refers to mapping only the strand in contact with protein (or hybridizing with RNA in case of R-loops). ***Directionality refers to mapping only the DNA strand involved in RNA-DNA hybrid formation and detecting a signal of strength representing the frequency of this DNA strand to form an RNA-DNA hybrid.*** It's noteworthy that strand specificity, but not high resolution, is a prerequisite for directionality but both don't confirm it.

Importantly, dsChIP-exo method can be used for general, robust and high-resolution mapping purposes such as studying signal localization over positive loci. However, some facts stand against directionality of dsChIP-exo, making it inconvenient to build a conclusion based on its results regarding directionality (Figure 4.28). The first fact is that the used sonication conditions produce average fragment size of ~200 bp (Figure 3.2 & Figure S1) while the fragment size of R-loops is variable, extending from few base pairs to few kilobases. This means short R-loops or, more precisely, R-loops shorter than average fragment size may stay intact after sonication as full R-loops with flanking dsDNA. On the other hand, R-loops that are longer than the average fragment size, will get sheared by sonication to smaller fragments (~200 bp). These smaller fragments resulting from shearing of long R-loops will lose the free ssDNA and produce only RNA-DNA hybrids. I's also important to consider that S9.6 captures the RNA-DNA hybrid not the R-loop *per se.*

The second fact is that dsChIP-exo relies on double stranded adaptor ligation. For this ligation to be successful, a dsDNA template is required for ligation to the adaptor. Simply, the first ligation wouldn't work if there were no dsDNA ends. Only short R-loops are expected to have dsDNA ends, i.e., even if the ssDNA is lost during sonication, these short R-loops would have dsDNA regions flanking the RNA-DNA hybrids, making them compatible for dsDNA adaptor ligation (Figure 4.28). Definitely, this is a limitation against the directionality principle because the second DNA strand may get captured and detected as a signal opposite to the main one even if there were no R-loop on the other direction. These short flanks may get recovered as short signals and confound the directionality.

On the other hand, long R-loops, or R-loops longer than the average fragment size will get sheared by sonication to produce RNA-DNA hybrids missing the free ssDNA and also dsDNA ends. It's not clear or understood how double stranded DNA adaptors can be ligated to the RNA-DNA hybrid. This implies that a second DNA strand may get generated during the library preparation, especially the polishing and phosphorylation steps. For these reasons,

directionality of dsChIP-exo method is doubted. This explains the difference in directionality between my dsChIP-exo and ssChIP-exo.

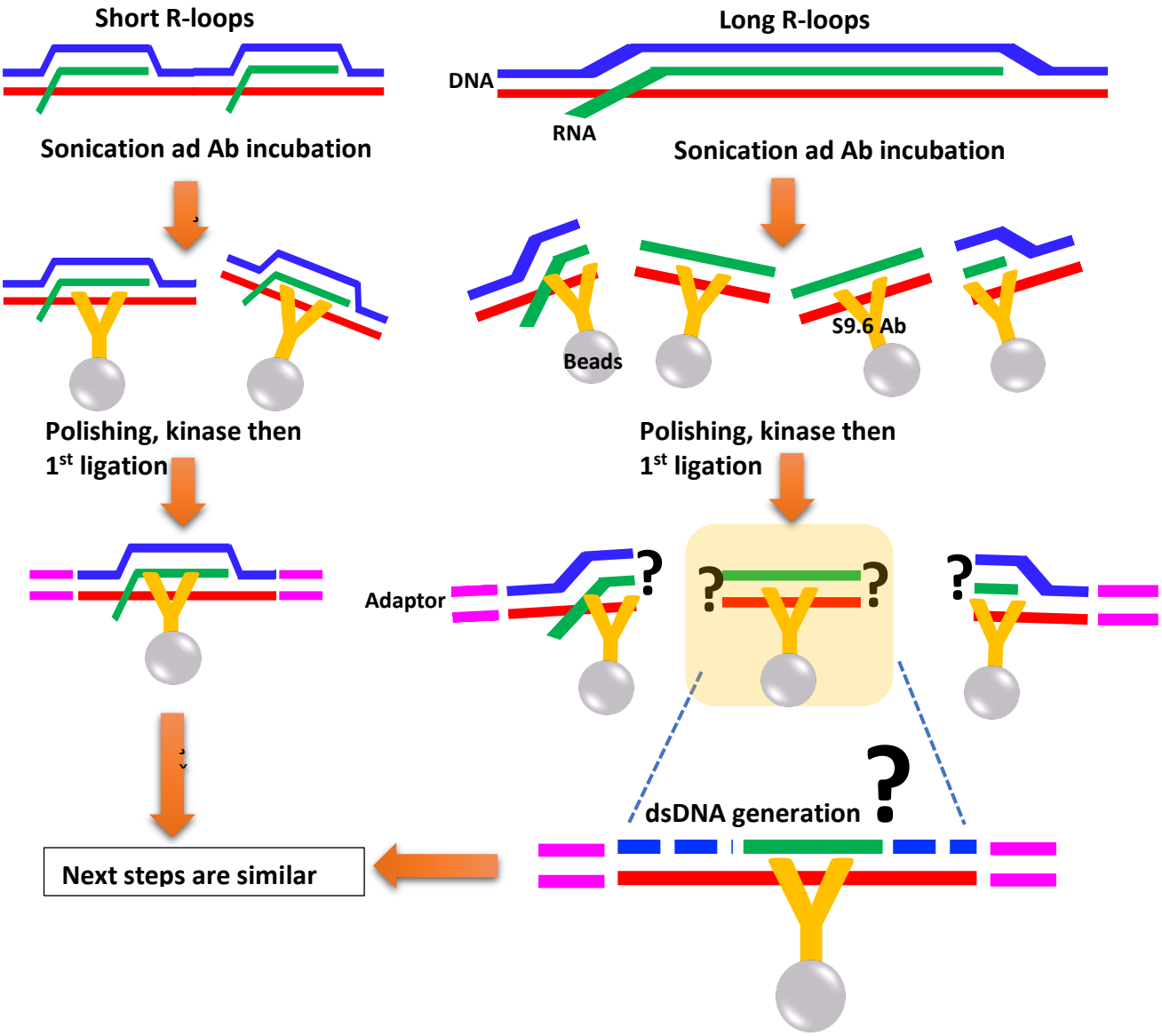



**Figure 4.28. Schematic representation for possible outcomes from sonication of short and long R-loops and their compatibility for ligation**. Short R-loops are expected to have dsDNA flanks after fragmentation. Long R-loops may produce RNA-DNA hybrids incompatible for ligation. It's not known whether polishing may generate canonical dsDNA which would affect directionality.

dsChIP-exo results regarding directionality of signals are reminiscent with those obtained by S1-DRIP which, also, depended on double-stranded adaptor ligation, but used S1 exonuclease to completely degrade the free ssDNA and prevent its rehybridization to the template DNA during sonication. Despite this, this strategy couldn't present directional mapping **(Wahba *et al.*, 2016)**. Moreover, using a spike-in RNA-DNA hybrid without free

ssDNA, but with flanking dsDNA ends, S1-DRIP produced a bidirectional signal that is typical for canonical dsDNA **(Wahba *et al.*, 2016)**. This suggests that a second DNA strand was generated during library preparation to represent a false opposite signal. For these reasons, a ssDNA-capturing method such as ssDRIP was introduced to enable the capture of only DNA strand of the hybrid and to deliver directional mapping **(Xu *et al.*, 2017)**.

The fact that ssChIP-exo produced more directional signals confirms that ssDNA is sheared and torn during sonication and cleared during washing steps, which is the same conclusion from two different studies using directional ssDNA-based methods, ssDRIP **(Xu *et al.*, 2017)** and qDRIP **(Crossley *et al.*, 2020)**. The fact that only short genes such as tRNAs show the highest degree of correlation between forward- and reverse-strand signals for dsChIP-exo and ssChIP-exo, suggests that ssDNA is lost during sonication of long R-loops, but not short R-loops. Retaining ssDNA of short R-loops may be the reason why these short genes produce double signals in both methods. This is supported by the fact that SMART-DRIPc produces single signals even over these short genes, except for very few ones.

This brings an important question about the source of the opposite signal in dsChIP-exo. Before the firs ligation step, there is a polishing step which is done to make sure both DNA strands are of the same length (blunt) and compatible for ligation to the same adaptor. dsDNA generation may happen through non-specific extension using small RNA or DNA fragments that can act as primers. If it happens that the original and newly synthesized DNA strands are ligated to the same adaptor, this means they will have very close chances to be captured and sequenced to produce two signals of comparable strength. However, one of the signals (template strand) was always much stronger. It's noteworthy that the same polishing step is done for ssChIP-exo, but the main difference form dsChIP-exo, is that polishing is followed directly by the exonuclease without ligation. The use of spike in and mimic R-loops would be helpful to understand how these steps work.

This situation for dsChIP-exo is reminiscent with directional RNA sequencing where the RNA is extended using random or oligo-dT primers to form an RNA-DNA hybrid which can't be captured by dsDNA adaptor ligation. To generate dsDNA, the RNA strand is nicked by RNase H to produce short primers that can generate the second DNA strand with the action of DNA Polymerase I (which is also used in dsChIP-exo and ssChIP-exo) and sometimes in the presence of T4 DNA ligase. Now the dsDNA can be ligated to forked adaptors which can still

produce strand specific, but not directional signals. To confirm directionality, the second DNA strand is labelled with dUTPs during the extension step to be digested later after ligation, and to keep only one strand with the required information for sequencing.

Importantly, the prerequisite for dsDNA ends may be a limitation against the sensitivity of the dsChIP-exo method, or methods depending on double stranded ligation, in general. For these reasons, a single-stranded DNA-dependent library preparation should be used instead.

### 4.4.6 SMART-DRIPc delivers directional signals and doesn't capture dsRNA

Different from **Hartono *et al.*, 2018**, I found that RNA-based R-loop mapping doesn't require special enzymatic treatments to provide stranded R-loop signals in fission yeast. Explicitly, SMART-DRIPc showed high resolution and strict directionality even without the use of RNase III or P1 (Figures 4.10, 4.12, 4.13 & 4.20). The main reason behind this difference may be the difference in immunoprecipitation substrate for DRIPc and SMART-DRIPc where nucleic acids extract was used for the former while chromatin was used for the later. Importantly, the RNIII/P1-treated sample showed lower number of signal peaks and weaker signals compared to non-treated sample, suggesting that these enzymes may have affected the efficiency of the mapping (Figures 4.10 & 4.13). A better experimental conduct would be to equilibrate these enzymatic reactions to define the most optimal working conditions. Moreover, it would be better to treat the samples with each enzyme separately.

### 4.4.7 Double R-loops may form over tRNA and rRNA genes

It's not interesting anymore that dsChIP-exo showed double signals genome-wide, which isn't supported by neither ssChIP-exo nor SMART-DRIPc (Figure 4.13). dsChIP-exo revealed a mild correlation between forward and reverse strand R-loops signals for *rnh1Δ rnh201Δ* cells, especially, for tDNA and telomeric and mitochondrial rDNA (Figure 4.16). This correlation disappeared with ssChIP-exo which showed a negative correlation for *rnh1Δ rnh201Δ* and WT cells (Figures 4.24 & 4.25). Consistent with anticorrelation between sense antisense transcription, SMART DRIPc showed a negative correlation between forward and reverse strand R-loops signals (Figure 4.17). Despite the revealed negative correlation, some of these regions still show R-loops signals over both DNA strands, as they may have high R-loops levels induced by high transcription rates. These few regions might be considered as potential double R-loop-forming regions. This observation is consistent with findings from mammalian

cells using R-ChIP which revealed that only tRNA genes accumulate R-loops signals on both sides **(L. Chen *et al.*, 2017)**.

It has been suggested that sequence homology and complementarity of tRNA sequences with both DNA strands allow tRNAs to anneal to both DNA strands, which may be the reason for detecting R-loops signals on both directions **(L. Chen *et al.*, 2017)**. If this has been the case, these double signals should have been revealed by both DNA-based (ssChIP-exo) and RNA-based (SMART-DRIPc) for all the regions and with the same ratio. Yet, this degree is variable among methods, so the detection of signals on both DNA strands may be dependent on the method used. Consistent with my observations, the RR-ChIP (RNA-based) produced signals more directional than those of R-ChIP (DNA-based) even over tRNA genes in mammalian cells. Nevertheless, my ssChIP-exo showed a difference between WT and *rnh1Δ rnh201Δ* cells in the degree of correlation between both DNA-strand signals, suggesting that this double signal may be real to some extent.

### 4.4.8 Most of R-loops are unstable against RNase H in WT cells

Comparing R-loops signals of WT and *rnh1Δ rnh201Δ* cells revealed different classes of R-loops based on sensitivity to physiological levels of RNase H. The first class represents RNase H-sensitive R-loops which are depleted in WT compared to *rnh1Δ rnh201Δ* cells. These R-loops belong to ~60 - 70% of R-loop-forming genomic regions (Figure 4.18). Of this class, are mitochondrial R-loops which tend to form on both DNA-strands. The second class represents RNase H-insensitive R-loops which don't decrease in WT compared to *rnh1Δ rnh201Δ* cells. Remarkably, in addition to rRNA genes known to accumulate RNase H-insensitive R-loops, tRNA genes accumulated strong R-loops in WT yeast cells (Figure 4.19).

In general, my observations are consistent with a previous study on fission yeast which identified RNase H-insensitive R-loops over 25% of forming regions. Some of these R-loops were still detected even after expressing bacterial RnhA in WT fission yeast **(Hartono *et al.*, 2018)**. My observations are also consistent with previous findings using quantitative DRIP (qDRIP) which identified RNase H-resistant R-loops *in vivo* in mammalian cells **(Crossley *et al.*, 2020)**.

Remarkably, all my methods detected very strong R-loops over tRNA genes in WT fission yeast. On the contrary, **Hartono *et al.*, 2018**, using DRIPc, detected R-loops over tRNA

genes only in *rnh1Δ rnh201Δ,* but not in WT fission yeast cells, and classified tDNA R-loops under the RNase H-sensitive category. My explanation for this discrepancy is that using phenol chloroform extraction and ethanol precipitation before immunoprecipitation in case of DRIPc may have denatured these R-loops, especially in WT cells. ssChIP-exo revealed that RNase H deletion increased the length of these R-loops (width of the signals) (Figures 4.12, 4.13 & 4.19) which may have enhanced their stability. The increased length and enhanced stability of tDNA R-loops with RNase H deletion may explain why DRIPc captured them only in *rnh1Δ rnh201Δ*, but not WT cells. This means more native conditions (chromatin preparation) are required to capture these short R-loops. The fact that neither DRIP nor DRIPc methods **(Sanz *et al.*, 2016),** but only R-ChIP detected R-loops over tRNAs in mammalian cells **(L. Chen *et al.*, 2017)**, strongly supports my explanation. DRIP and DRIPc methods use nucleic acids extract for the IP while R-ChIP uses chromatin preparation for mapping dynamic R-loops using more native conditions **(L. Chen *et al.*, 2017)**.

Importantly, sensitivity of intergenic R-loops which extend beyond tDNA borders toward RNase H in WT cells, suggests that RNase H may target these regions mainly to digest and trim the extended extragenic R-loops, but not the intragenic ones (Figures 4.12, 4.13 & 4.19). Probably, another enzyme or an RNA-DNA helicase may be in charge of processing the intragenic hybrids at these regions. This suggestion is based on the finding that Sen1 mutation increases R-loops levels mainly over short genes in budding yeast **(Chan *et al.*, 2014)**. In addition, Sen1 was found to be specifically enriched at these short RNA Pol III-transcribed genes, in fission yeast, compared to the rest of the genome **(Rivosecchi *et al.*, 2019)**. However, the fact that RNase H can recognize and bind tDNA R-loops in mammalian cells doesn't support this suggestion. These R-loops were captured by the RNase H-based R-ChIP method **(L. Chen *et al.*, 2017)**. Nevertheless, it should be considerd that the binding of RNase H enzyme to an R-loop sequence deosn't confirm the enzymatic activity of this enzyme over this sequence. Probably, recognition, binding and processing of R-loops represent distinct functions. Also, R-loops may not bind or process all R-loops at an equal rate **(Crossley *et al.*, 2020)**.

### 4.4.9 Some RNase H-insensitive R-loops overlap with asRNA and dsRNA signals

It makes sense to conclude that RNase H-sensitive R-loops (unstable R-loops) can't participate in dsRNA formation as they are actively degraded. An exception for this category,

is mitochondrial R-loops which form heavily on both DNA strands (Figure 4.21), making it plausible for double R-loop or double R-loop-dependent dsRNA formation over mitochondria. On the other hand, the RNase H-insensitive R-loop category (stable R-loops) may be suggested to form dsRNA. However, most of these R-loops form only on one DNA strand, which doesn't support double R-loop or double R-loop-dependent dsRNA formation. An exception for this category is telomeric rDNA and some tDNA R-loops which form on both DNA strands (Figure 4.19), not excluding dsRNA formation through double R-loops over telomeric rRNA genes.

Interestingly, I found the stable single R-loops to be associated with sense antisense transcription sites, and importantly, to overlap with dsRNA signals, suggesting a correlation between stable R-loops and dsRNA formation. One big dilemma regarding this finding is that some of these dsRNA-forming regions didn't show clear antisense mature or nascent transcription (Figure 4.20). It's noteworthy that **Hartono *et al*., 2018** reported that RNase H manipulation didn't affect the level of these dsRNAs which means they weren't engaged in R-loop formation before forming dsRNA. Anyway, none of the dsRNAs showed a response to RNase H manipulation, but only to the exosome (**Hartono *et al*., 2018).** A more dedicated investigation and dsRNA-mapping method (instead of DRIPc which captured these dsRNAs) is required.

### 4.4.10 Challenges for R-loop-dependent dsRNA formation: R-loop-asRNA is more plausible than double R-loop model

As explained above, dsRNAs have been shown to form over the genome for many organims. In fission yeast, dsRNAs have been enriched over 35% of fission yeast genome during a trial to map R-loops using DRIPc **(Hartono *et al.*, 2018)**. These dsRNA structures were confirmed by their sensitivity to RNase III, the dsRNA-splicing enzyme. This may suggest that there is a correlation between R-loops and these dsRNAs. However, this doesn't seem to be the case, as the dsRNAs weren't sensitive to *in vivo* RNase H overexpression or to *in vitro* RNase H treatment. Moreover, different from R-loops which form mainly over gene bodies, these dsRNAs were enriched over intergenic regions. Finally, these structures were sensitive to the exosome and accumulated upon deletion of Rrp6 subunit of the nuclear exosome. These last two facts suggest that these dsRNA-forming transcripts belong to a set of transcripts different from R-loop forming transcripts.

Regardless of the dependence of these dsRNAs on R-loops, it couldn't be explained how these dsRNAs are formed given the known global weak to negative correlation between sense antisense transcription. This global anti-correlation under physiological conditions between sense and asRNA transcription at mature RNA level **(Atkinson *et al.*, 2018)** and at the nascent transcript level **(Murray *et al.*, 2015; Wery, Gautier, Descrimes, Yoda, Vennin-Rendos, *et al.*, 2018)** represent a big challenge for R-loop-mediated dsRNA formation as well. It has been revealed that genes with high transcription rates will have very low to no asRNA transcription **(Atkinson *et al.*, 2018)**. In concordance, antisense transcripts predominantly accumulate over overlapping protein coding genes with low sense transcription levels. Furthermore, the more the overlap between the sense and antisense transcript, the stronger the anticorrelation. This anticorrelation reaches its maximum when the antisense transcript overlaps with the TSS of the sense one **(Wery, Gautier, Descrimes, Yoda, Vennin-Rendos, *et al.*, 2018)**. These findings suggest that most of R-loop-forming genes which are highly transcribed, will have a low antisense transcription level and, accordingly, a low chance for double R-loop and dsRNA formation.

Besides, negative correlation was also shown to exist between R-loops and antisense transcription. A previous genome wide study examined the correlation between R-loops and antisense transcription at R-loop forming loci using *E. coli* cells deficient for rho termination factor known to accumulate R-loops. This study found that degrading R-loops by expressing UvsW, a viral helicase, elevated antisense transcription levels over 500 R-loops-forming loci. This suggests a negative impact for R-loops on antisense transcription **(Raghunathan *et al.*, 2018)**. These findings support my observations for anticorrelation between forward- and reverse-strand R-loops over tDNA and rDNA in WT (ssChIP-exo) and *rnh1 rnh201 ΔΔ* cells (ssChIP-exo & SMART-DRIPc) (Figures 4.16, 4.17, 4.24 & 4.25). This is also consistent with my observation that an increase of R-loops signals over one DNA strand decreases the other strand signals, and that overaccumulation of R-loops signals on one strand abolishes the signal of the other (Section 4.3.14).

A recent finding that sounds contradicting at first glance, but needs careful interpretation, is that R-loops have been shown to act as Pol II promoters and enhance antisense transcription in mammalian cells. Importantly, these antisense RNAs formed proximal to the R-loops and didn't strictly overlap with the RNA-DNA hybrid **(Tan-Wong *et al.*, 2019)**. In essence, this finding doesn't clash with the anticorrelation established above, instead, it shows that sense-

antisense RNAs don't overlap which means there is no chance for dsRNA formation at R-loop forming loci.

As explained in text, R-loop-dependent dsRNA formation is mainly supported by the fact that different forms of ssDNA are transcriptionaly competent and can also form RNA-DNA hybrids, e.g., over DSB ends **(Ohle *et al.*, 2016)**. However, the relevance and reproducibility of these results for free ssDNA in an R-loop context need to be confirmed. It's hard to speculate how ssDNA of an R-loop would behave compared to ssDNA gap or overhang at the transcription initiation level. Moreover, R-loops are known to block transcription elongation *in vitro* **(Belotserkovskii & Hanawalt, 2011**, **2015; Belotserkovskii *et al.*, 2010; Belotserkovskii, Mirkin*, et al.*, 2013; Belotserkovskii, Neil*, et al.*, 2013; Belotserkovskii *et al.*, 2017; Tous & Aguilera, 2007)** and *in vivo* **(González-Aguilera *et al.*, 2008; Huertas & Aguilera, 2003)**. Finally, ssDNA over DSB can form a stable hybrid that's not threatened by the presence of a second complementary DNA strand, but in R-loops, there is always a thermodynamic competition between displaced ssDNA and RNA strand of the hybrid to anneal with template DNA strand and restore dsDNA.

Other facts stand against double R-loop model, in particular. It's known that RNA-DNA hybrids form over purine (GA) rich or high GC-skew DNA where the RNA-DNA hybrid formation requires specific features such as the presence of Guanine repeats in RNA strand for the threading back and initiation of R-loop formation. This means while the template DNA strand and sense RNA will have a high propensity to anneal and form an RNA-DNA hybrid, the non-template strand and antisense RNA will not have the same propensity as the antisense RNA will be pyrimidine rich. This is supported by the fact that nucleosome occupancy decreases over a bidirectionally transcribed DNA when G- but not C-rich RNA is generated from the same DNA **(Powell *et al.*, 2013)**. In addition, different ssDNA-binding proteins tend to bind ssDNA on a sequence-recognition basis. This binding stabilizes the R-loop and, at the same time, may block transcription over the ssDNA **(Cristini *et al.*, 2018)**. A clear example for these proteins, is AtNDX protein which binds ssDNA at the flowering locus in *Arabidopsis* **(Sun *et al.*, 2013)**. For these reasons, double R-loop formation seems to be a weaker model compared to the R-loop-antisense RNA model.

It seems the second model, suggesting dsRNA formation through an RNA-DNA hybrid overlapping with free asRNA, is more plausible than double R-loop model. This is supported

by the observation that stable R-loops exist over sites of sense-antisense transcription and dsRNA formation. Still a major limitation against this model (and the first model as well) is the weak correlation between nascent sense and antisense RNA which means that both RNA transcripts have to be transcribed at different temporal intervals. However, this would decrease the chance of hybridization between both transcripts. As a solution, one of RNA strands needs to wait in vicinity (at the local transcription site) till the transcription of the second strand. Interestingly, it has been shown that some RNAs stay intact and associated with chromatin **(Bronner *et al.*, 2017; Tan-Wong *et al.*, 2019)**.

Regardless of these obstacles and challenges, double R-loops have been detected in some biological systems, but as very rare events. Using DRIPc, **Sanz *et al.* (2016)** reported a small fraction of genomic regions with R-loops forming in the sense and antisense direction in mammalian cells. Interestingly, they detected double R-loops only at converging terminators with short intergenic distances implying that double R-loops formation is still limited to specific conditions **(Sanz *et al.*, 2016)**. Here, I detected double R-loops signals over only tRNA and telomeric and mitochondrial rRNA genes in WT cells using ssChIP-exo.

### 4.4.11 RNase H deletion may induce global transcriptional changes and impact R-loop levels on both DNA strands in a contrasting way

The decrease for R-loop levels in WT compared to *rnh1Δ rnh201Δ* cells, or more precisely, the increase of R-loop levels in *rnh1Δ rnh201Δ* compared to WT cells can be explained by sensitivity of these R-loops to RNase H. However, it's hard to explain the reverse situation where RNase H deletion decreased R-loop levels over template strand of some sense genes. This was revealed by ssChIP-exo for R-loops over ribosome biogenesis and proteins synthesis genes (Figures 4.19 & 4.22). From the RNase H sensitivity perspective, this implies insensitivity of these R-loops to RNase H and absence of RNase H local enzymatic activity at these genes. Importantly, this suggests a decrease for the transcription rate and a change in expression level of these genes. My explanation is based on the fact that transcription is the main determinant for R-loop formation. This explanation is also supported by the fact that RNase H deletion decreases the growth rate of the cells, referring to a change in the transcriptomic and proteomic profiles of these cells. Importantly, genes showing this effect are transcribed in the sense direction, and their R-loops belong to the RNase H-insensitive category which form on reverse DNA strand (Figures 4.19 & 4.22).

In agreement with my conclusion (which need to be confirmed by RNA-seq experiments), **Hartono *et al.*, 2018** suggested that long term manipulation of RNase H, affects the transcriptomics and proteomics of fission yeast cells. They found that RNase H deletion or overexpression in fission yeast affects transcription of ribosome biogenesis and protein synthesis genes. However, and in contrast, they reported that these two different conditions upregulate the same genes. Explicitly, upregulation of gene expression doesn't match a decrease in R-loop levels over template strand of sense ribosome biogenesis genes since this contradicts with known correlation between R-loop formation and transcription rate.

It's noteworthy that RNase H deletion induced a contrasting effect on antisense-transcribed compared to sense ribosome biogenesis and protein synthesis genes (Figures 4.18 & 4.23). RNase H deletion increased R-loop levels over template strand of these genes. This may match gene upregulation reported by **Hartonon *et al.*, 2018**. These antisense genes are characterized by RNase H sensitive R-loops forming on forward strand. Closer inspection showed that RNase H deletion seems to have genome wide and contrasting impacts on R-loops of both DNA strands. In essence, RNase H deletion increased R-loops over forward DNA strand and decreased them over reverse strand compared to wild type cells. While this suggests RNase H sensitivity of most forward strand R-loops, but not reverse strand ones, this also refers to a decrease of transcription over the reverse strand upon RNase H deletion.

Assuming that changes in R-loop levels represent transcriptional changes, these observations mean that RNase H manipulation may affect expression of different genes based on their transcription direction, i.e., RNase H deletion will decrease transcription of sense-transcribed genes and, on the other hand, increase transcription of antisense-transcribed genes. If true, this would suggest a role for RNase H in suppressing forward strand R-loop formation and maintaining transcription of sense-transcribed genes. This may explain why RNase H deletion and overexpression are thought to produce similar effects. My results tend to agree with **Chan *et al.* 2014** who found that overexpression of RNase H significantly upregulated sense transcription of R-loop forming genes of sense-antisense transcription in budding yeast.

This fluctuation in R-loop levels may reflect changes in DNA replication and differences in leading- and lagging- strand generation as the latter requires more hybrids formed by Okazaki RNA fragments. Importantly, these results may need to be explained in connection with DNA replication as well. My observations are reminiscent with recent findings combining

R-loops, DNA replication, RNA transcription and DNA damage response. DNA damage response pathway was found to be dependent on the orientation of transcription-replication collision **(S. Hamperl *et al.*, 2017)**. Possibly, depletion of RNase H may accumulate R-loops that would first block transcription and also stall replication and suppress cellular activities. This may explain why RNase H depleted cells grow slower than WT cells. *rnh1Δ rnh201Δ* may have stalled replication, paused transcription and decelerated protein synthesis processes which, by their turn, can decrease transcription as a feedback mechanism. The decrease in the levels of R-loops that are genuinely insensitive to RNase H, over some genes, may reflect this decrease in transcription in *rnh1Δ rnh201Δ* cells.

Notably, the cause-effect correlation for RNase H deletion effect on R-loops is still not clear. It can't be confirmed whether the reverse-strand R-loop levels decrease as a result of increase in forward-strand R-loop levels due to deletion of RNase H. Also, it can't be excluded that RNase H deletion may have induced transcriptional changes (decrease) for sense transcription which may have enhanced antisense transcription and, accordingly, forward strand R-loop levels. Consequently, it needs to be confirmed whether the increased forward-strand R-loops in RNase H-depleted cells match an increase of transcription on this strand or whether this only occurs due to RNase H deletion. The fact that the forward strand signal is high genome-wide even at non-R-loop forming regions doesn't support the explanation that this is a direct effect for local R-loop levels changes (Figure 4.18, check extensive region).

Finally, a great benefit for these observations, is that they confirm that the enhanced signals over the RNase H-depleted cells, especially, over the forward strand, and the opposite response for the reverse strand, aren't outcomes for PCR bias or sequencing artefacts. I believe these observations regarding differential enrichment of signals in *rnh1Δ rnh201Δ* compared to WT are enough to validate the results better than the use of RNase H treatment control. It's very hard to find a region with similar R-loop levels in WT and *rnh1Δ rnh201Δ* to use as internal control for normalization. Therefore, the use of spike-in R-loops as a standard for quantifying R-loops and comparing global R-loop enrichment in *rnh1Δ rnh201Δ* compared to WT cells would be critical to confirm these results.

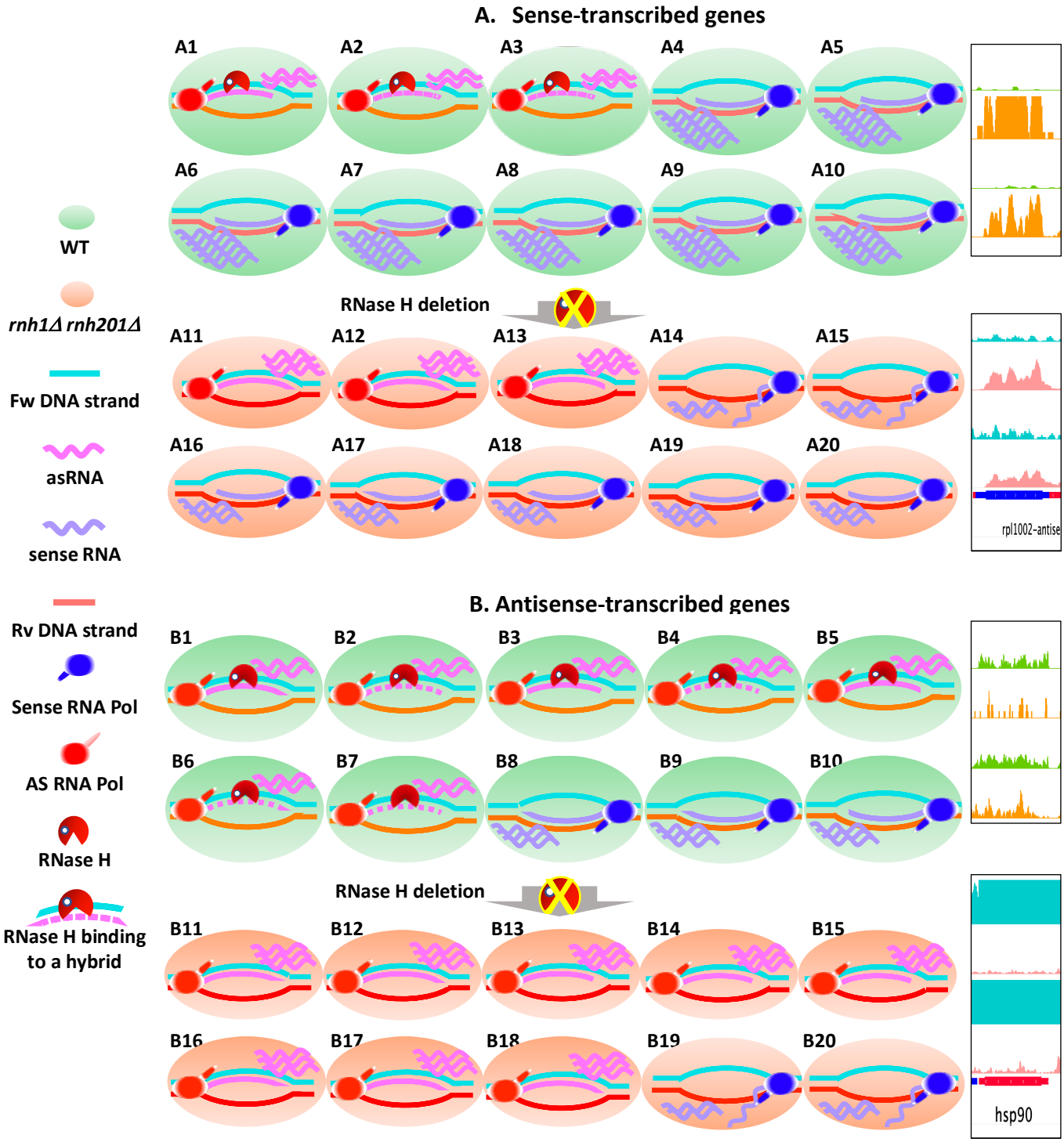



**Figure 4.29. RNase H deletion increases forward-strand R-loops while decreases reverse-strand R-loops. (A)** sense genes with few forward-strand R-loops and abundant RNase H-insensitive reverse-strand R-loops: RNase H binds and resolves forward-strand hybrids (A1-A3), but not intragenic reverse strand hybrids (A4-A10): RNase H deletion allows formation of forward strand hybrids which may be associated with increase of antisense transcription (A11-A13): This may decrease sense transcription (A14-A20) and reverse-strand hybrids (A14-A15). **(B)** antisense genes with abundant but RNase H-sensitive forward-strand R-loops and rare reverse-strand R-loops: RNase H binds and resolves forward-strand hybrids (B1-B7), but not reverse strand hybrids (B8-B10): RNase H deletion induces accumulation of forward-strand R-loops (B11-B18) and abolishes reverse-strand R-loops (B19, B20). These two processes are expected to be associated with a change in levels of sense and antisense transcription.

## 4.5 Conclusion

The election for R-loop mapping technique and experimental conditions to be applied, is very critical for consistency of results, as inappropriate consideration for conditions can affect the whole results and conclusions.

It seems to be very hard to confirm a role for R-loops in dsRNA generation or even to confirm double R-loop formation. Detection of clear signals on both DNA strands doesn't confirm formation of double R-loops, as it needs to be shown that these signals belong to the same sequence of one single cell. For this, signle cell R-loop mapping is required. Despite being challenging, this would be the most definitive approach. The same situation applies to R-loops and antisense transcription knowing that multiple findings refer to a negative correlation between R-loops and antisense transcription.

Given the fact that different factors such as capturing methods, experimental conditions and cell types may contribute to discrepancies in R-loop mapping results, it would be important to confirm my results and check their reproducibility using another S9.6-independent method, such as RNase H based methods for example. A suggested change is to adopt a method that uses more *in situ* (*in vivo*) conditions to decrease the effect of lysis and sonication on stability of R-loops.

The directionality of any R-loop mapping method must be verified and validated, especially in case of studying double R-loop formation and for strand-specific applications. This can be achieved using artificial R-loop controls of known sequences and of known R-loop formation direction. Likewise, the performance of known R-loop mapping and library preparation methods including current ones toward mapping and sequencing double R-loops should be tested. By other words, the competency of current methods for mapping double R-loops need to be confirmed as some factors might affect their potential for capturing or sequencing double R-loops. For example, the strong binding of RNA to DNA at some sequences may affect annealing of extension primer or hinder strand synthesis and amplification rounds, especially at RNase H resistant sites with high GC content and high melting temperatues. This may affect detection and mapping of some R-loops in general and double R-loops in particular. However, the fact is R-loops are detected over GC rich sequences and correlate with GC content. These sequences are known to have strong annealing and high melting temperature, which weakens the possibility that these factors

(strong annealing) may prevent capturing or sequencing of double R-loops. Anyway, this needs to be practically tested to exclude such concerns. Overall, it can't be completely excluded that double R-loops may exist at specific regions.

The best available tool for manipulating R-loops levels to study their role based on cellular response to their depletion or accumulation, is through manipulation of RNase H levels. However, RNase H manipulation seems to induce a global change in transcriptional and proteomic landscape of cells, independent from R-loops. This makes it practically very hard to understand a correlation between R-loops, antisense transcription and dsRNA formation.

Here, I show that RNase H deletion may affect transcription of ribosome biogenesis, protein synthesis and translation related structural tRNA genes, suggesting a global response that may be independent form R-loops. My observations in the next chapter that RNase H-deletion depleted K3K9me2 signals over all R-loop-free heterochromatic regions support this suggestion.

I also show that RNase H deletion increases R-loops levels on forward strand while decreases their levels on reverse strand. This opposing effect for RNase H deletion refers to a role for RNase H in regulating sense-antisense transcription. More excitingly, it suggests a difference in R-loop regulation by (response to) RNase H that is based on R-loop orientation/formation direction. Finally, and independently, this suggests a role for RNase H in regulating gene expression based on transcription direction of affected genes. For this role RNase H suppresses forward-strand R-loops to maintain the expression of reverse strand (sense genes). For this, further investigation is required to decipher the main mechanism behind this response toward RNase H depletion, especially the change of signal over both strands. Does it correlate with transcriptional changes over both strands? is it just a direct response to increase of signals on forward strand following an anticorrelation pattern? is it dependent on R-loops or due to a deficiency of RNase H?

# Chapter 5. Investigating Link Between R-loops, Termination and Heterochromatin Formation

## 5.1 Introduction

R-loops have been suggested to play a role in transcription termination and heterochromatin formation. This has been supported by the fact that R-loops signals were detected over terminator regions in budding yeast **(Wahba *et al.*, 2016)**, in fission yeast under genetic conditions **(Castel *et al.*, 2014; Ohle *et al.*, 2016)** and in mammalian cells **(Ginno *et al.*, 2013; Nadel *et al.*, 2015; Sanz *et al.*, 2016).** Moreover, R-loops were reported to form over pericentromeric and heterochromatic repeats in fission yeast **(Castel *et al.*, 2014; Nakama *et al.*, 2012)**, in budding yeast under genetic and pathological conditions **(Mishra *et al.*, 2021)**, in mammalian cells **(Nadel *et al.*, 2015)** and in *Arabidopsis* **(Xu *et al.*, 2017)**.

Intriguingly, RNase H overexpression which resolves R-loops, was found to perturb transcription termination and lead to transcriptional readthrough beyond terminal pause sites of some mammalian genes **(Skourti-Stathaki *et al.*, 2011)**. Moreover, RNase H1 overexpression was found to abolish dsRNA formation, deplete RNAi machinery proteins such as Dcr1, Ago1, G9a histone lysine methyltransferase, H3K9me2 histone mark, and, importantly, heterochromatin protein 1 (HP1) over the same constructs **(Skourti-Stathaki *et al.*, 2014)**. These two studies from the same lab strongly supported a role for R-loops in transcription termination and heterochromatin formation. However, both studies involved either engineered transfected genic constructs or few endogenous genomic loci. Accordingly, a genome-wide role for R-loops in such processes still need to be confirmed.

Among genomic regions that has been reported to be enriched by R-loops are telomeric repeats in budding yeast **(Chan *et al.*, 2014; El Hage *et al.*, 2014; Wahba *et al.*, 2016),** mammalian cells **(Nadel *et al.*, 2015; Rosso & d'Adda di Fagagna, 2020)** and *Arabidopsis* **(Xu *et al.*, 2017)**. Telomeres represent important heterochromatic regions in some organisms such as fission yeast. If true, formation of R-loops over telomeres would suggest a role for R-loops in heterochromatin formation. Independently, R-loops have been reported to be involved

in maintaining telomeres and telomer homeostasis in different systems **(Bettin *et al.*, 2019; Niehrs & Luke, 2020; Santos-Pereira & Aguilera, 2015)**.

The fission yeast *S. pombe* is a powerful model organism for studying heterochromatin formation and the underlying mechanisms, especially the RNAi pathway. In *S. pombe*, heterochromatin exists over specific genomic loci which include peri-centromeric repeats and the associated tDNA clusters, mating loci, telomeric DNA repeats and ribosomal DNA. Histone H3 di-methylated at lysine 9 residue (H3K9me2) is a hallmark of heterochromatin at these regions. This histone methylation mark is added by Clr4 enzyme, the key writer (modifying enzyme) responsible for this modification. Heterochromatin assembly requires the RNA interference (RNAi) machinery which, in brief, starts with the action of RNA directed RNA polymerase complex (RDRC) that amplifies Pol II-transcribed long noncoding RNA at these regions to generate dsRNA. dsRNA molecules are spliced by Dicer 1 (Dcr1), an RNase III-family enzyme, into small interference RNA (siRNA) molecules that get loaded onto the argonaute protein Ago1, a member of the argonaute complex (ARC). This loading elicits the assembly of the RNA interference transcriptional silencing complex (RITS) which later recruits the cryptic loci regulator complex (CLRC) containing Clr4 that writes the methylation mark. H3K9me2 triggers the docking of Swi6, a homolog of the heterochromatin proteins (HP1), required for the silencing **(Allshire & Ekwall, 2015; Martienssen & Moazed, 2015)**.

## 5.2 Hypothesis and aims

Based on previous lab data and recent findings from other labs, I speculated that R-loops may have a role in heterochromatin assembly and transcription termination. In fact, I had a broader and more comprehensive hypothesis for R-loops as potential nuclear regulators for gene expression. Broadly, in retrospect, I expected that R-loops may act as progenitors for dsRNA formation and that the resultant dsRNAs may represent precursors for small interference siRNA and nucleate heterochromatin assembly in fission yeast through the RNAi machinery. However, since I couldn't assure an R-loop-dependent dsRNA formation, I sought to independently investigate a plausible role for R-loops in heterochromatin assembly and transcription termination on a genome-wide scale, mainly through studying the abundance of R-loops at terminators and heterochromatic regions of *S. pombe.*

Exploiting the high resolution and directional R-loop mapping data of this study, I investigated the differential enrichment of R-loops over different sub-genic compartments,

especially terminators. I also studied the localization of R-loops over heterochromatic domains in fission yeast. To test whether R-loops contribute to heterochromatin assembly, I examined the result of RNase H depletion in fission yeast on the enrichment of H3K9me2 histone mark. Moreover, I examined the result of RNase H overexpression in U2OS mammalian cells on the same mark. Unexpectedly, I didn't detect R-loops at any of the heterochromatic regions except tRNA clusters at the pericentromeric region and rDNA repeats at telomeres of chromosome III. R-loops were confined to intragenic regions and uniformly distributed over gene bodies without special enrichment at gene ends. Strikingly, RNase H depletion in fission yeast severely depleted H3K9me2 at heterochromatic regions. Independently, RNase H overexpression in U2OS mammalian cells disrupted H3K9me2 genome-wide.

## 5.3 Results

### 5.3.1 R-loops are restricted to intragenic regions and uniformly distributed over gene bodies

Bearing in mind that R-loops were reported to be enriched at gene terminators, I sought to verify these observations using my own R-loop mapping data generated by the three different methods. In general, and consistent with recent findings from *S. pombe* using DRIPc **(Hartono *et al.*, 2018)** and *Arabidopsis* using ssDRIP **(Xu *et al.*, 2017)**, R-loops signals were restricted to intragenic regions and uniformly distributed over gene bodies of transcribed regions. As may have been noticed in all figures of the previous chapter, there was no bias for R-loop signals to peak over 5' or 3' ends of genes. Only dsChIP-exo for *rnh1Δ rnh201Δ* cells detected rare loci with intergenic R-loop peaks around termination region. One of these R-loop-forming loci belonged to a short intergenic region between converging terminators of two colliding genes while the second one showed R-loops over terminator regions of overlapping sense-antisense transcribing genes (Figure 5.1). These observation matches the major theme identified for mammalian cells using DRIPc **(Sanz *et al.*, 2016)** which found that terminal R-loops form at convergent terminators with short intergenic regions. Anyway, metagenic analysis for dsChIP-exo data didn't support terminal R-loop formation as metagenic signal was depleted over termination region (Figure 5.1). Metagenic analysis for ssChIP-exo data of non-crosslinked WT cells revealed a slight increase in R-loops reads count around transcription start site. On the contrary, the same analysis revealed a drop in R-loops reads count at transcription end site in agreement with external DRIPc (+RNase III) data for WT **(Hartono *et al.*, 2018)** that was included in the same analysis (Figure 5.1). These rare examples for terminal R-loops detected

by only one method (dsChIP-exo), don't support R-loop formation at terminators in fission yeast.

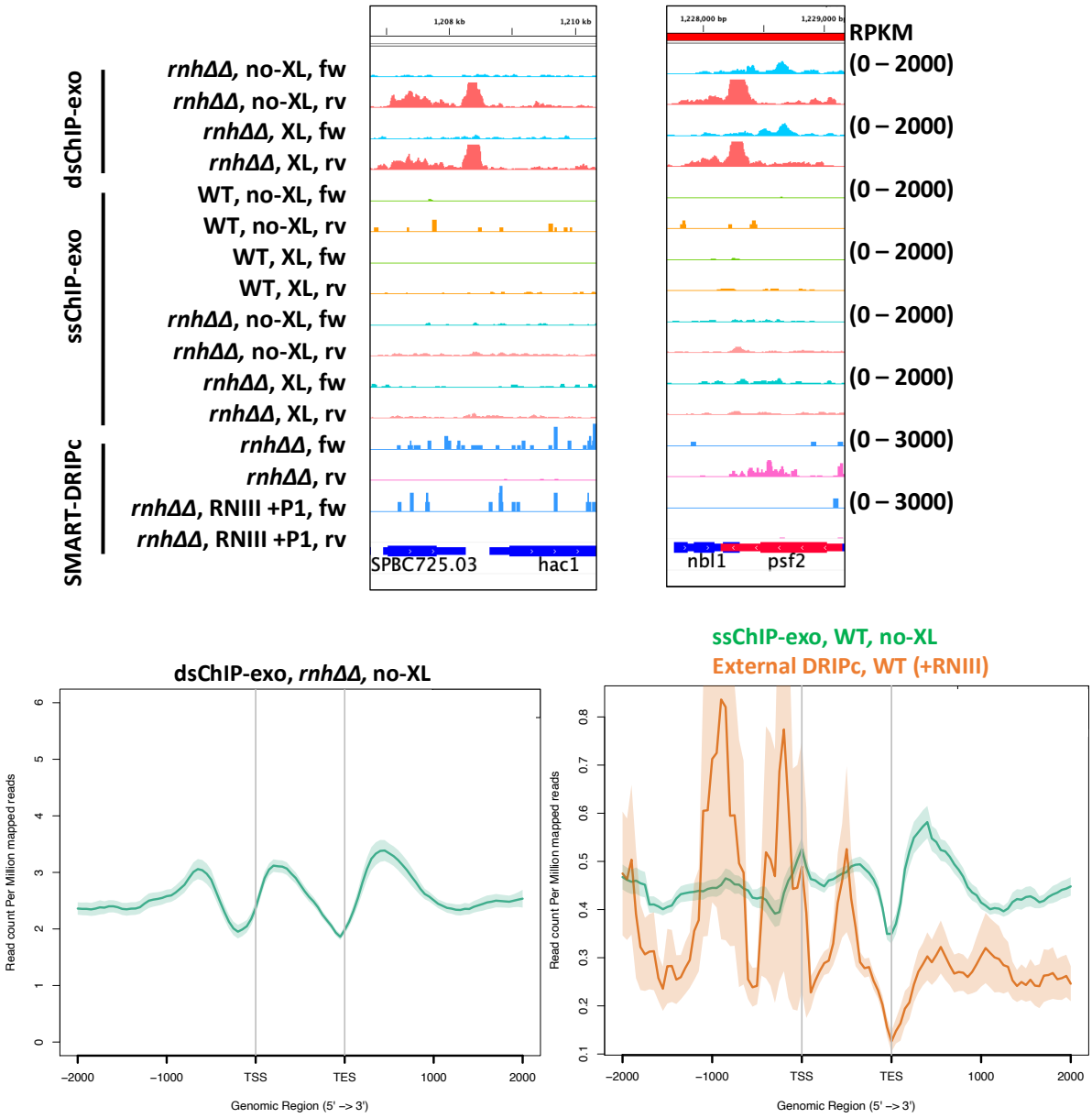



**Figure 5.1. R-loops signals are depleted over gene terminators in fission yeast.** Top, examples for two regions showing terminal R-loops detected only by dsChIP-exo. The two regions belong to convergent terminators of colliding genes or overlapping sense-antisense genes. Bottom, Metaplots showing reads count decreases over transcription end site (TES). External DRIPc data for RNIII-treated WT (red) was included in the analysis.

Apparently, it has been reported that R-loops form over terminator regions of fission yeast. However, I found that this was shown for Dicer 1-deleted (*dcr1Δ*), but not WT cells **(Castel *et al.*, 2014)**. For budding yeast, amongst multiple reports, only S1-DRIP detected R-

loop peaks over terminator regions **(Wahba *et al.*, 2016)** suggesting a specific bias for this method. This implies that the difference in genotype (in case of fission yeast) or methodology (in case of budding yeast) may be the reasons for this discrepancy. However, this explanation doesn't fit for a previous publication mapping R-loops for *rnh1Δ rnh201Δ S. pombe* cells and using the same method (ChIP-exo) **(Ohle *et al.*, 2016)**. This was very hard to explain at the beginning, but after careful inspection, I realized that the public data exhibited a shift of signals to either the promoter or terminator regions. This pattern wasn't produced when the data was reanalyzed. These findings suggest that the algorithm used may be the reason for this inconsistency.

### 5.3.2 R-loops don't form over constitutive heterochromatin in fission yeast

Next, I investigated the enrichment of R-loops at constitutive heterochromatic regions in fission yeast. These include pericentromeric repeats, telomers, mating loci and tRNA clusters. Strikingly and as stated in previous chapter, all my methods showed very strong signals over the tDNA clusters surrounding the pericentromeric repeats in all chromosomes. However, such signals weren't detected for the pericentromeric repeats. In fact, I couldn't detect clear R-loops signals over neither pericentromeric repeats nor the mating loci (Figure 5.2). Instead, weak patchy signals were produced using ssChIP-exo over pericentromeric repeats of WT, but not *rnh1Δ rnh201Δ* cells (Figure 5.2).

These profiles specific for my data are completely different from the profiles of multiple external DRIP data which showed very strong and extensive signals covering the tRNA genes and the pericentromeric repeats **(Bronner *et al.*, 2017; Castel *et al.*, 2014; Ohle *et al.*, 2016)**. Nevertheless, my findings are consistent with those of **Hartono *et al.*, 2018** who didn't detect R-loops over fission yeast heterochromatin using DRIPc.

Notably, I detected a similar situation for telomeric repeats in fission yeast. While telomeric repeats of chromosome III were densely packed with R-loops signals which were the strongest signals all over the genome, telomeric repeats of chromosome I and II didn't show obvious signals (Figure 5.2), except for tlh2 gene at sub-telomeric region of chromosome II. Remarkably, reanalyzed external DRIPc data of **Hartono *et al.*, 2018** supported my observation. Importantly, telomeric repeats of chromosome III host the most highly-transcribed

18S and 28S rRNA genes. These observations suggest that R-loops are marks of active transcription rather than silent heterochromatin.

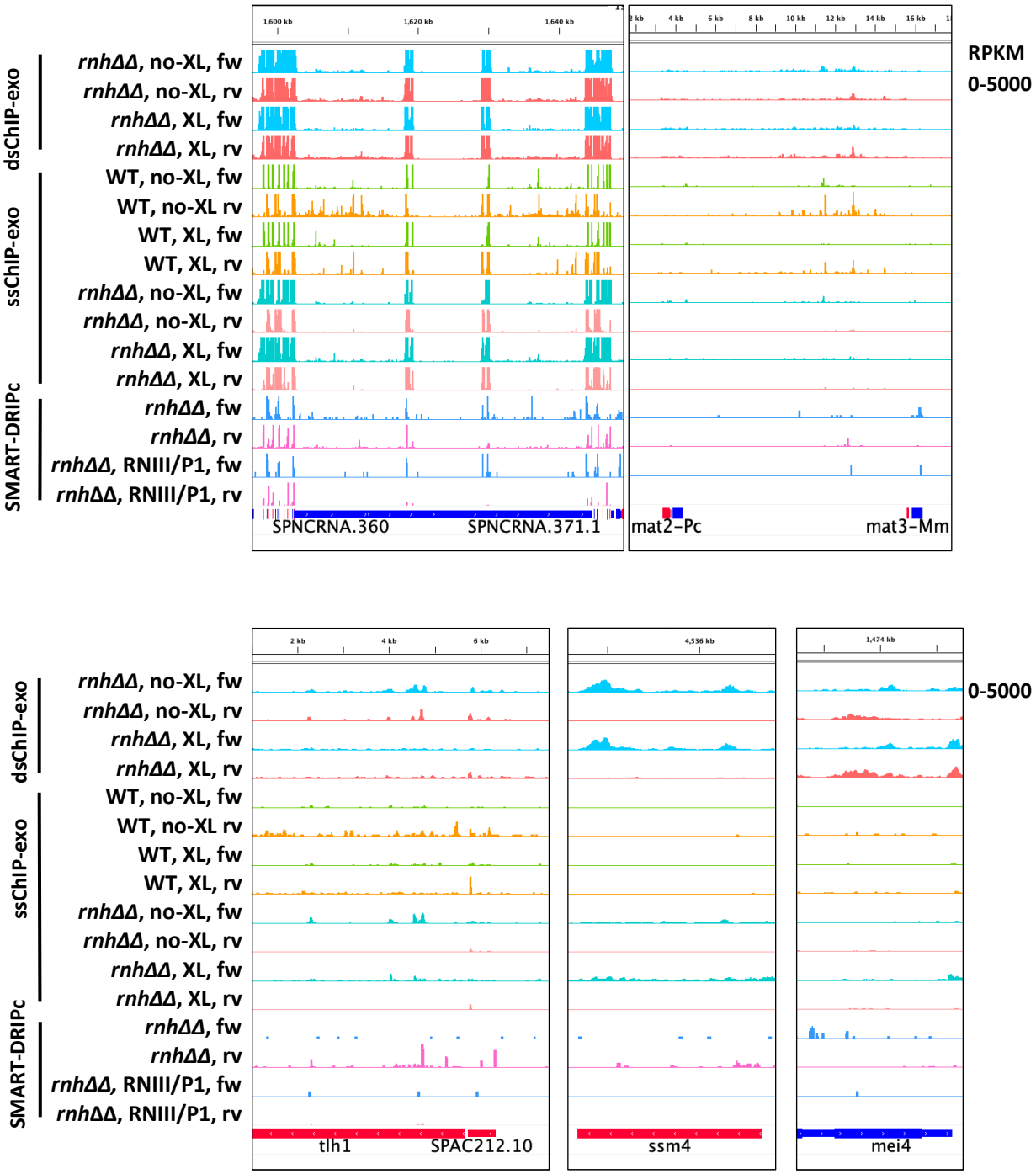



**Figure 5.2. No R-loop signals are detected over constitutive or facultative heterochromatin in fission yeast.** Snap shot of genome browser showing R-loops forming over tDNA clusters but not over pericentromeric repeats (top left) mating loci (top right), telomeres (bottom left) or facultative heterochromatin (ssm4 and mei4).

Furthermore, I tried to investigate R-loop formation over previously identified facultative heterochromatic islands in fission yeast **(Zofall *et al.*, 2012)**. Again, my methods didn't reveal clear R-loops signals over these islands (Figure 5.2). My observations do not support the formation of R-loops over neither constitutive nor facultative heterochromatin in fission yeast, which, accordingly, weakens the plausibility for a role played by R-loops at these regions.

### 5.3.3 H3K9me2 signals are associated with R-loops signals at some loci

Next, I moved to check whether any correlation exists between R-loops and H3K9me2 histone mark outside the canonical heterochromatic regions of fission yeast. To address this, I mapped H3K9me2 mark using the ChIP-exo-TT technique for the WT and *rnh1*Δ *rnh201*Δ yeast cells (Chapter 3). After cell lysis, sonication and immunoprecipitation, bead-bound chromatin was treated directly with T7 exonuclease before elution and DNA purification. Oligo-dT primer was annealed to DNA for extension and dsDNA generation before final ligation and single end sequencing. Next, I validated the data using standard PCR, exploiting the fact that the strains have two copies of ura4 inserted at different genomic region, one full length ura4 inserted at pericentromeric region, and another truncated ura4 inserted at euchromatic region. Indeed, standard PCR produced a single signal specific for full heterochromatic ura4 but not the truncated one. The no-Ab was empty of any amplification (Chapter 3, Figure 3.7).

As expected, the strongest signals were detected over pericentromeric repeats, mating loci, all telomeric repeats and tDNA clusters. In addition, strong signals were detected over meiotic genes such as mei4, ssm4 and mcp7 (Figures 5.3 & 5.5), representing facultative heterochromatic islands in fission yeast **(Zofall *et al.*, 2012)**. Surprisingly, ChIP-exo-TT signals strongly enriched over background, were detected over multiple genic regions representing previously unknown H3K9me2 foci. Some of signals over these foci were more presented compared to those over previously identified heterochromatic islands (Figure 5.3, compare to ssm4 and mcp7). These islands belonged to different regions from the three chromosomes: Chr1; pfl2, prl53, mae1, rpl302, not3, pyk1, fas1, hsp90 and pma1: Chr2; mfs3, tef103, sks2, but2, snou14, fba1, tdh1, mbx2 and sam1: and Chr3; htb1/sec5, SPCC1281.06c.1, adh1, tef3 and ssa2. Interestingly, most of these regions are associated with either convergent or overlapping sense antisense transcription (Figures 5.3 & 5.5).

Next, I checked for a correlation between H3K9me2 and R-loops signals at these regions. Intriguingly, I observed that H3K9me2 signals coincide with R-loops signals over these regions, supporting a correlation between these two factors. Examples for genes with such correlation are Pyk1, pma1, hsp90, tef103, sks2, but2 (Figure 5.4). Some of these regions were shown to form strong R-loops (Chapter 4).

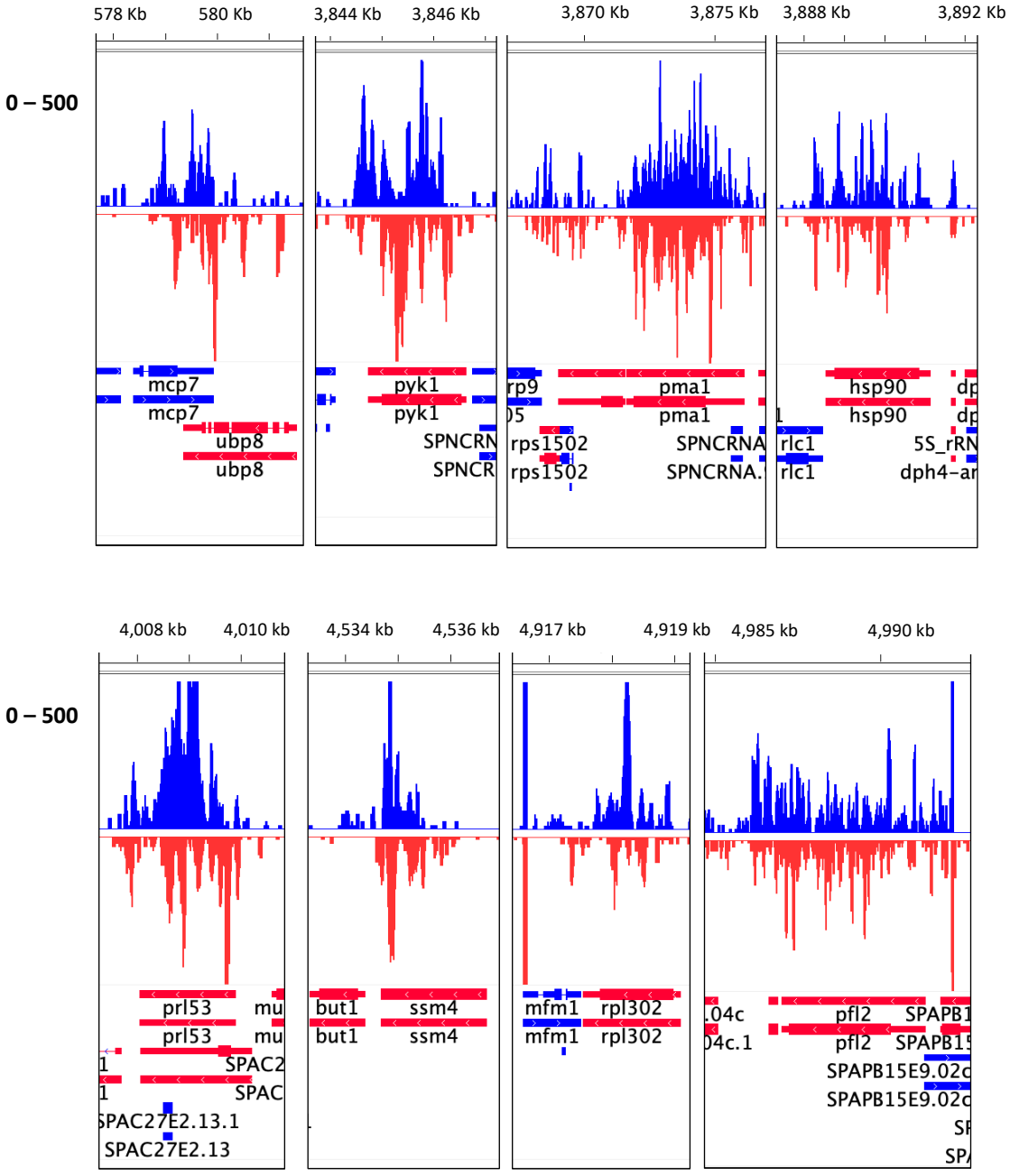



**Figure 5.3. Examples for previously unidentified H3K9me2 foci in fission yeast.** Snap shot of genome browser, showing regions that are more enriched with H3K9me2 signals compared to ssm4 and mcp7 meiotic genes. Signals were detected with ChIP-exo-TT method. Examples are shown for chromosome 1 only.

I have previously shown that R-loops are associated with regions of antisense transcription and also coincides with dsRNA signals (Chapter 4, Figure 4.20), so I checked whether these R-loop forming regions coincide with H3K9me2 signals, but didn't detect H3K9me2 signals over them.

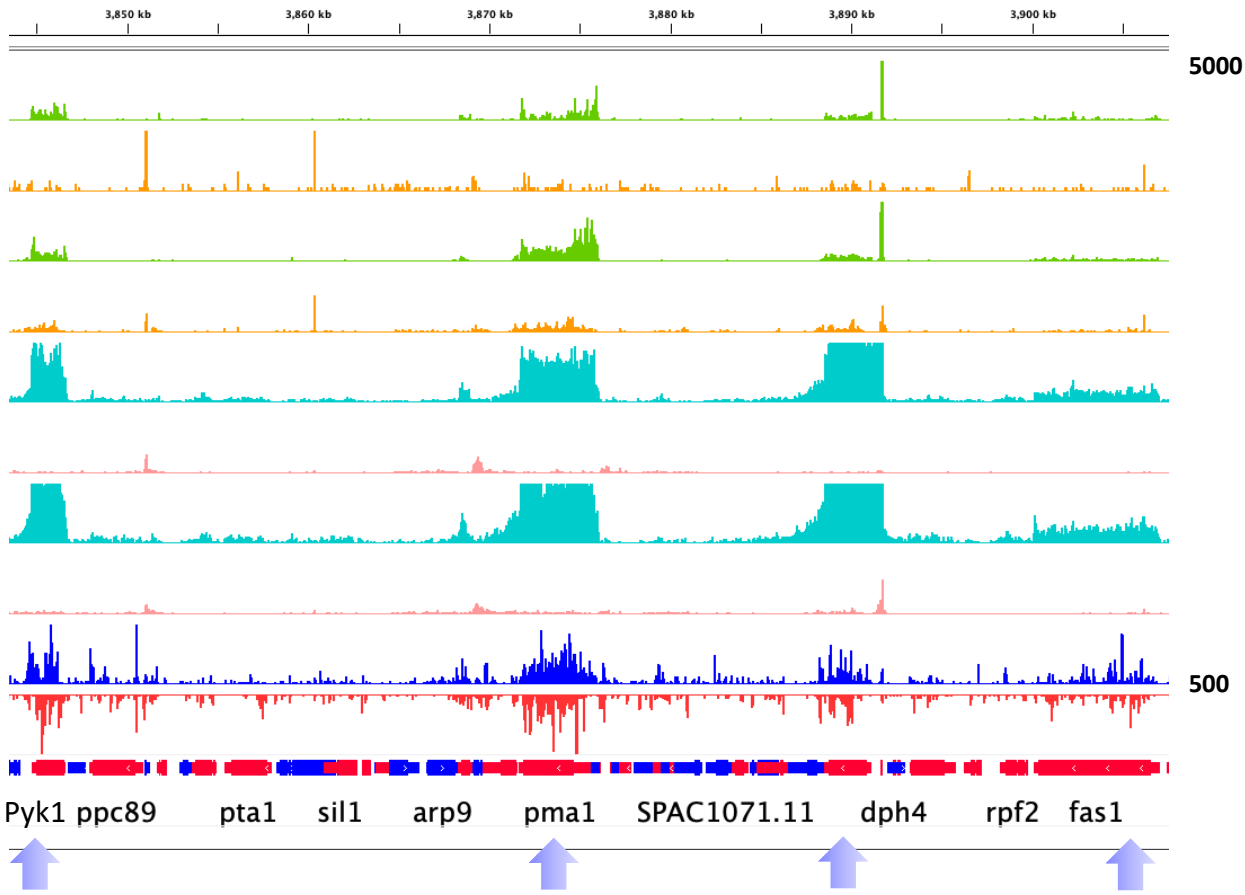



**Figure 5.4. H3K9me2 signals are associated with R-loops signals over previously unidentified H3K9me2 islands.** Arrows point to regions of association.

### 5.3.4 RNase H deletion severely depleted H3K9me2 over heterochromatin in fission yeast

Although R-loops weren't detected over known constitutive or facultative heterochromatic regions, it was interesting to see that some R-loops signals are associated with H3K9me2 signals over some genes that may represent novel facultative heterochromatic foci. This motivated me to check whether any correlation exists between R-loops and H3K9me2 signals over these regions. If a correlation exists between R-loops and H3K9me2, H3K9me2 enrichment is supposed to change according to change in R-loops levels. By another meaning, RNase H depletion which accumulates R-loops should increase H3K9me2 levels compared to

WT. To test this possibility, I depended on my H3K9me2 mapping data for WT and *rnh1Δ rnh201Δ* cells. Unexpectedly, RNase H deletion tended to deplete H3K9me2 signals over these regions, suggesting a negative effect for RNase H depletion on heterochromatin formation (Figure 5.5).

It has been previously reported that R-loops accumulation disrupts H3K9me2 mark over pericentromeric and telomeric repeats in fission yeast **(Bronner *et al.*, 2017; Nakama *et al.*, 2012)**. Despite the fact that I didn't detect R-loops over these regions, I sought to examine the effect of RNase H deletion on H3K9me2 mark over these regions using my own data. Explicitly, H3K9me2 signals over these heterochromatic regions were severely depleted in *rnh1*Δ *rnh201*Δ compared to WT cells (Figure 5.5).

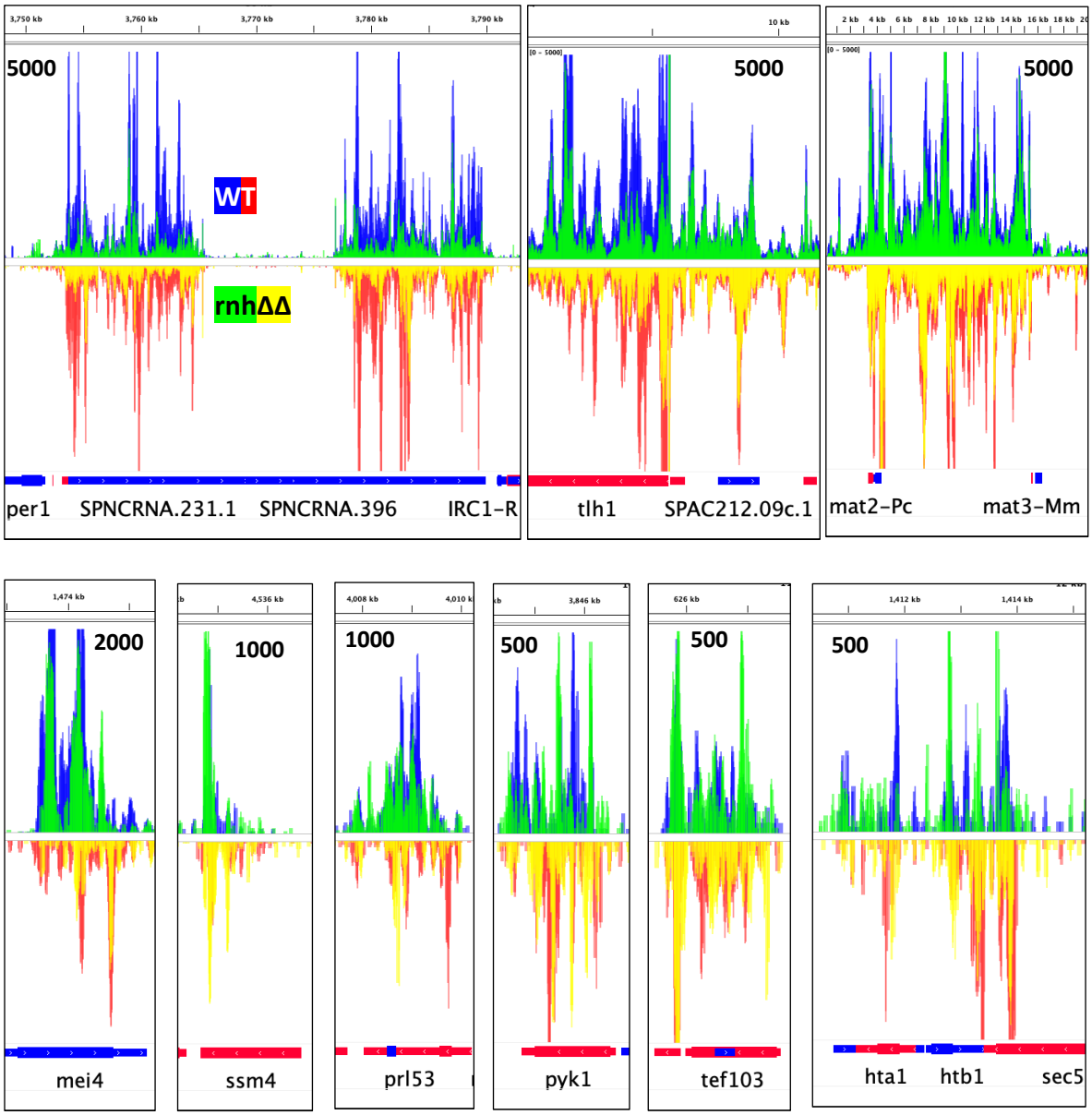



**Figure 5.5. RNase H deletion depletes H3K9me2 signals over heterochromatin in fission yeast.**

It doesn't make sense to think that a local change in R-loops levels at these heterochromatin regions, is the reason for H3K9me2 signal depletion since these regions are devoid of R-loops. Thus, this can't be attributed to impairing H3K9me2 writing machinery. Neither may it be an impact on nucleosome deposition. A better explanation is that RNase H depletion affects the transcriptomics and proteomics of cells which tend to have lower transcription and translation rates and protein levels which affect cellular activities and growth rate.

### 5.3.5 RNase H overexpression in mammalian cells depleted H3K9me2 signals genome wide

The above results, obtained for fission yeast, suggest that the opposite may be true, i.e., overexpressing RNase H which resolves R-loops may increase H3K9me2 signals. This time, I used mammalian U2OS cells to check this possibility and, at the same time, to test the effect on another biological system. I carried out the immunoprecipitation for WT and RNase H1 overexpression cells and validated the results by comparing the produced profiles to those of another external data. Unexpectedly, H3K9me2 signals of RNase H-overexpression cells were depleted genome wide compared to WT cells (Figure 5.6).

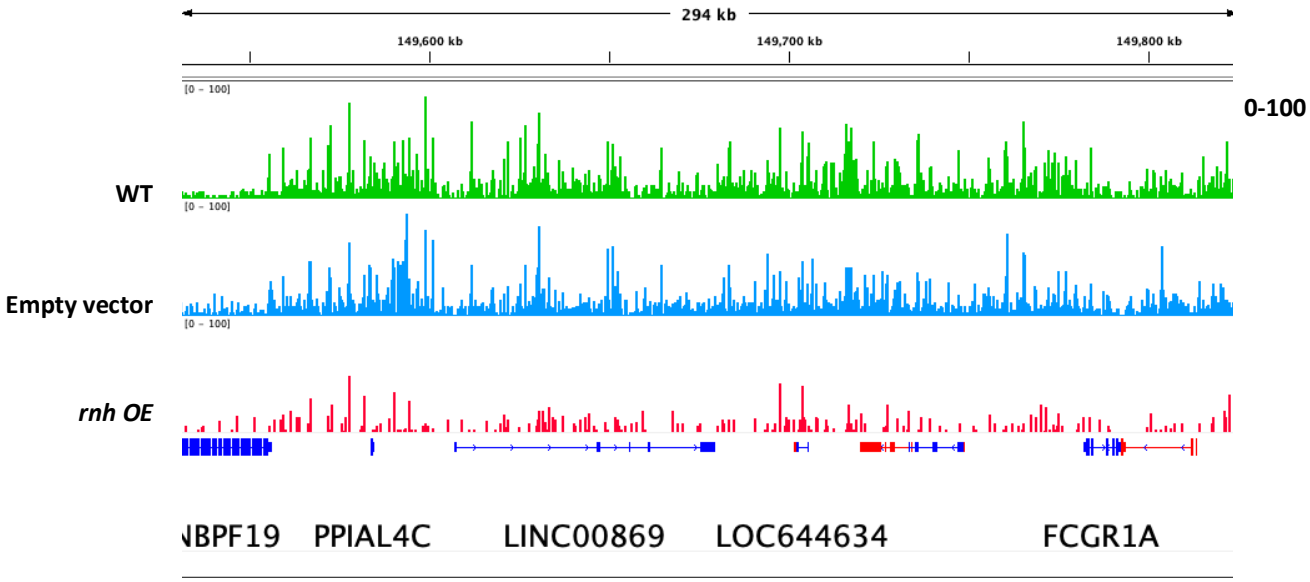



**Figure 5.6. RNase H overexpression depleted global H3K9me2 signals in U2OS cells.**

## 5.4 Discussion

### 5.4.1 R-loops and termination

Independent form dsRNA, I asked whether R-loops may have a role in transcription termination and heterochromatin formation. To answer this question, I used my powerful R-loop profiling data to see whether R-loops form over terminators and, separately, heterochromatin in fission yeast. None of my methods revealed a tendency for R-loops signals to peak over 5' or 3' ends of genes. Moreover, the signals over terminators were weaker than those over gene bodies. Explicitly, R-loop signals sharply declined downstream the 3' UTR (Figure 5.1). Only dsChIP-exo, but not ssChIP-exo or SMART-DRIPc, detected terminal R-loops signals over very scarce regions (Figure 5.1). Consistent with my observations, **Hartono *et al.*, 2018** haven't detected R-loops over terminators of fission yeast using DRIPc. Instead, R-loops were depleted over terminators, a pattern similar to that captured by ssDRIP for *Arabidopsis* **(Xu *et al.*, 2017)**.

The reason for the clash between my results and other reports for fission yeast, in my opinion, is the lack of resolution for most of the methods used in these studies. The fact that my findings are consistent with those of **Hartono *et al.*, 2018** who didn't detect R-loops at these regions in WT *S. pombe* cells using the high resolution DRIPc, strongly supports this explanation. However, it may be true that R-loops form over terminators under genetic conditions, which was revealed by DRIP for *dcr1Δ* **(Castel *et al.*, 2014)** and by ChIP-exo for *rnh1Δ rnh201Δ S. pombe* cells **(Ohle *et al.*, 2016)**. Nevertheless, reanalyzing the data for the later didn't reveal such features, suggesting a bias induced by the algorithm or analysis pipelines which induced a shift in signals toward 3' end of genes. The situation isn't much different for budding yeast. Re-evaluating the situation for budding yeast, revealed that only S1-DRIP data detected R-loops peaks at gene terminators in *S. cerevisiae* **(Wahba *et al.*, 2016)**, suggesting a bias for this method. Obviously and in contrary to some previous findings, my data do not support a role for R-loops in termination at least in fission yeast cells.

The global uniform distribution of R-loops along gene bodies in fission yeast seems to be different from mammalian cells which have been confirmed to accumulate R-loops over non-methylated CGI promoters and CG rich terminators of densely-packed genomic regions **(Chedin & Benham, 2020; Ginno *et al.*, 2013; Ginno *et al.*, 2012; Nadel *et al.*, 2015; K. Yu *et al.*, 2003)**. The absence of CGI promoters in fission yeast may explain why R-loops don't

form over these regions in fission yeast. However, the reason behind the difference regarding terminator regions isn't clear. While it can be explained by the difference in the CG content between the two organisms, a mechanistic reason may lie behind this difference. Probably, a long R-loop-forming sequence may be required for R-loops to exert an effect on Pol II termination. Terminator region is quite broad in mammalian cells **(Sanz *et al.*, 2016)** reaching 10 – 20 kb which is longer than the average length of *S. pombe* genes. It has been shown that the formation and position of R-loops around the poly-Adenylation signal (pAs) of gene terminators in mammalian cells affect poly-adenylation factor PAF1 levels and localization, and, accordingly, Pol II release point and termination **(Sanz *et al.*, 2016)**.

A strong evidence for a role of R-loops in transcription termination has been presented by Proudfoot lab who found that RNase H overexpression, which resolves R-loops, perturbs transcription termination and leads to transcriptional readthrough beyond terminal pause sites of transfected genic constructs **(Skourti-Stathaki *et al.*, 2011)**. They also suggested that these terminal R-loops are required for recognition by Sen1 to recruit the 5' – 3' exonuclease Xrn2/Rat1 which cleaves the 3' transcripts downstream the co-transcriptional termination and cleavage (CoTC) site, thereby facilitating pol II termination and release. Although this refers to a role for R-loops and Sen1 in transcription termination, the conservation of this role is questionable. Recently, Sen1 has been found to be required for transcription termination at RNA Pol III-transcribed genes in fission yeast. Depletion of Sen1 led to accumulation of RNA Pol III downstream these genes and produced 3' extended transcripts. However, this process involving Sen1 was independent of R-loops formation **(Rivosecchi *et al.*, 2019).** In this regard, although it was found that depletion of Sen1 or DXH9 accumulates R-loops **(Cristini *et al.*, 2018; Skourti-Stathaki *et al.*, 2011)**, the former was found to be unable to resolve RNA-DNA hybrid *in vitro* **(Porrua & Libri, 2013)**.

Importantly, even detection of R-loops at termination regions in mammalian cells is dependent on methodology used. While DRIPc and DRIP **(Sanz *et al.*, 2016)** supported the enrichment of R-loops at terminators and their association with termination features, RDIP **(Nadel *et al.*, 2015)**, R-ChIP **(L. Chen *et al.*, 2017)** and RR-ChIP **(Tan-Wong *et al.*, 2019)** didn't agree with these findings. Instead, these methods revealed that R-loops are depleted around transcription end site in mammalian cells. Notably, while Proudfoot lab suggested a role for R-loops in termination **(Skourti-Stathaki *et al.*, 2011)**, they didn't detect clear R-loop

signals peaks over terminators compared to promoters in mammalian cells using RR-ChIP **(Tan-Wong *et al.*, 2019)**.

My results along with recent findings suggest that R-loop formation over terminator regions may be restricted to mammalian cells. R-loops formation over terminators in mammalian cells may be predisposed by specific nucleotide and chromatin features **(Ginno *et al.*, 2013; Ginno *et al.*, 2012; Sanz *et al.*, 2016)** that may not exist in fission yeast. The fact that R-loops form over terminators in mammalian cells, but not in cells of other organisms, implies that R-loops may not contribute directly to transcription termination, but, rather indirectly, through the action of other molecules. This also suggests differences in underlying mechanisms regulating transcription termination in different organisms. In addition, a direct role for R-loops in transcription termination may be context dependent and can't be generalized. NET-seq experiments could be conducted to provide more clear insights about Pol II termination and kinetics in relation to R-loops, through monitoring nascent transcription in response to RNase H deletion or overexpression.

**Table 5.1. List of different methods used for mapping R-loops over promoters, terminators, telomeres and heterochromatin.** Table indicates existence (+) or absence (-) of R-loop signals over these regions, and highlights discrepancy among findings for different organisms. NA, not applicable; * form reanalysed public data.

| System | Reference | Method | Genotype | Promoter | Terminator | Centromeric | Telomeric |
|---|---|---|---|---|---|---|---|
| **Mammalian** | **(Ginno *et al.*, 2013; Ginno *et al.*, 2012)** | DRIP | WT | + | + | NA | NA |
| | **(Sanz *et al.*, 2016)** | DRIPc | WT | + | + | - | NA |
| | **(L. Chen *et al.*, 2017)** | R-ChIP | WT | + | - | NA | NA |
| | **(Tan-Wong *et al.*, 2019)** | RR-ChIP | WT | + | - | NA | NA |
| | **(Nadel *et al.*, 2015)** | RDIP | WT | + | - | + Chr. 9 | + |
| | | | | | | | |
| ***Arabidopsis*** | **(Xu *et al.*, 2017)** | ssDRIP | WT | + | - | + | + |
| | | | | | | | |
| **Fission yeast** | **(Castel *et al.*, 2014)** | DRIP | WT<br>*dcr1Δ* | NA<br>NA | -<br>+ | +<br>+ | NA<br>NA |
| | **(Bronner *et al.*, 2017)** | DRIP | WT<br>*caf1Δ*<br>*dcr1Δ* | NA<br>NA | NA<br>NA | +<br>+ | -<br>+ |
| | **(Ohle *et al.*, 2016)** | ChIP-exo | *rnh1Δ*<br>*rnh201Δ* | NA | + | + | +* |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **(Hartono *et al.*, 2018)** | DRIPc | WT<br>*rnh1Δ*<br>*rnh201Δ* | -<br>- | -<br>- | -<br>- | -*<br>- |
| | | | | | | | |
| **Budding yeast** | **(El Hage *et al.*, 2014)** | ChIP-DRIP | WT | NA | NA | NA | + |
| | **(Chan *et al.*, 2014)** | DRIP-microarray | WT | NA | NA | NA | + |
| | **(Wahba *et al.*, 2016)** | S1-DRIP | WT | + | + | - | + |

Although a role for R-loops in telomere maintenance has been established for some biological systems such as budding yeast and mammals **(Balk *et al.*, 2013; Bettin *et al.*, 2019; Perez-Martinez *et al.*, 2020),** my current results suggest that fission yeast may diverge in dependence on R-loops for such a role. None of my methods detected R-loops signals over telomeric sequence away from rDNA of chromosome III telomeres. My methods profiles for fission yeast are completely different from profiles of other systems showing significant and strong R-loop formation over telomeres, in budding yeast for example **(Chan *et al.*, 2014)**. A role for R-loops in telomere maintenance in fission yeast has been recently studied, but this requires careful consideration **(Hu *et al.*, 2019)**.

### 5.4.2 R-loops and heterochromatin

A spectacular problem with studying R-loop formation over pericentromeric repeats of fission yeast is that these regions are flanked by the highly expressed tRNA genes. Without enough resolution and in the presence of long DNA fragments, a long signal covering tRNA genes and extending over pericentromeric repeats will be always shown by R-loop mapping methods. This is exactly the situation for external DRIP methods **(Bronner *et al.*, 2017; Castel *et al.*, 2014; Ohle *et al.*, 2016)** (Figure 4.12). However, and in concordance with my results (Figure 5.2), **Hartono *et al.*, 2018** couldn't confirm the presence of R-loops over these regions using the high resolution DRIPc method.

Independently, in WT budding yeast, there is no support for R-loops formation over centromeres **(Costantino & Koshland, 2018; Mishra *et al.*, 2021; Wahba *et al.*, 2016)**. No wonder that their accumulation at these regions, under genetic conditions, affects kinetochore integrity and contributes to chromosomal instability **(Mishra *et al.*, 2021)**. For mammalian cells, only RDIP detected R-loops at only chromosome 9 centromeres. Likewise, only ssDRIP detected R-loops at centromeres in *Arabidopsis* **(Xu *et al.*, 2017)**. Genome-wide mapping

methods used to detect R-loops in different organisms over different genomic regions and genic compartments are listed in Table 5.1, showing variation among results of these methods.

While all the heterochromatic regions including telomeres didn't show distinguishable R-loop signals, only right and left telomeres of chromosome III were loaded with strong R-loops signals (Chapter 4). These signals delineate the ribosomal RNA genes, 18S, 25S and 5.8S, and mark them as the strongest R-loop-forming genes over the whole genome, even after accounting for the repetitive nature of these ribosomal RNA genes. This is quite reminiscent with the situation of tRNA genes forming at the boundaries and over the pericentromeric internal repeats where only tRNAs accumulate very strong signals. These observations support the fact that R-loops are marks of active transcription rather than heterochromatin determinants.

### 5.4.3 H3K9me2 signals overlap with R-loop signals over some regions

While my methods didn't detect R-loops over terminators or constitutive heterochromatin in fission yeast, I detected an association between H3K9me2 and R-loops signals over few and previously unidentified facultative heterochromatic islands (Figures 5.3 & 5.4). Identification of these islands/loci was enabled by the sensitive and high-resolution ChIP-exo-TT technique. Importantly, H3K9me2 signals detected by ChIP-exo-TT over these new loci, were stronger than those over previously known islands such as mcp7 and ssm4 (Figure 5.3). Similar to my findings for fission yeast, R-loops were found to be rarely associated with H3K9me2 repressive mark in *Arabidopsis* **(Xu *et al.*, 2017)**.

Perhaps, the strongest evidence ever supporting contribution of R-loops to heterochromatin formation and transcription termination was provided by Proudfoot lab who showed that R-loops are required for deposition of repressive H3K9me2. This was confirmed by the observation that RNase H overexpression depletes heterochromatin factors, H3K9me2 mark and heterochromatin protein 1 **(Skourti-Stathaki *et al.*, 2014)**. However, this was shown for pause elements and terminators of few genes. Importantly, the same lab found R-loops to be more associated with enhancer histone marks, but not heterochromatin **(Tan-Wong *et al.*, 2019)**.

Although R-loops were found to form over gene termini and to correlate with H3K9me2 in mammalian cells, the same R-loop-forming termini were found to be DNase hyper-sensitive

and to have accessible chromatin features, but not heterochromatic traits **(Sanz *et al.*, 2016)**. These findings suggest that even with a proposed role for R-loops in termination in mammalian cells, this role may be independent from heterochromatin. These findings also confirm that R-loops are more correlated with open and active euchromatin, but not silenced heterochromatin. Consistent with this, I detected an association between R-loops and only mild H3K9me2 signals, but not strong ones (Figures 5.3 & 5.4). It's noteworthy that these regions were always associated with sense-antisense transcription which would be interesting to investigate for a correlation between these three features, R-loops, H3K9me2 and antisense transcription (Figures 5.3 & 5.5).

### 5.4.4 RNase H manipulation may impact transcriptional and proteomic cellular profiles

Here, I show that in fission yeast, RNase H deletion which accumulates R-loops, depleted H3K9me2 histone mark over the pericentromeric, telomeric and mating loci heterochromatic repeats (Figure 5.5). My findings are in concordance with previous studies on fission yeast **(Bronner *et al.*, 2017; Nakama *et al.*, 2012)** which showed that accumulation of RNA over chromatin disrupts heterochromatin. However, these two studies attributed this effect to local accumulation of R-loops at these regions. This conclusion is mainly supported by detection of R-loops over heterochromatin. However, the fact that none of my methods nor DRIPc **(Hartono *et al.*, 2018)** could detect R-loops over these regions, makes it hard to explain this effect by local R-loop accumulation. This difference in conclusion can be explained by differnece in resolution of methods used for R-loop mapping. Decrease of H3K9me2 in RNase H depleted yeast cells over heterochromatic repeats, may be due to an indirect effect on cellular activities, especially transcription and translation which are required for steady-state supply of proteins and histones. In other words, RNase H depleted cells may have lower transcription and protein synthesis rates and lower proteome content. This effect for RNase H deletion is reminiscent with decrease of R-loop signals over specific genes in RNase H-depleted compared to WT cells which is also consistent with the fact that affected genes belonged to ribosome biogenesis and protein synthesis genes (Chapter 4). Importantly, I explained this effect by a decrease in transcription rate over these genes.

Unexpectedly, RNase H overexpression depleted the global H3K9me2 in U2OS mammalian cells (Figure 5.6). Although it may be irrelevant to draw a general conclusion based on the similar results from fission yeast, these observations suggest that either RNase H

deletion (global R-loops accumulation) or overexpression (global R-loop resolution) affects cellular homeostasis. This conclusion is consistent with the previous finding that either R-loops depletion or accumulation affects transcription termination and gene regulation **(Skourti-Stathaki *et al.*, 2011)**. My conclusion is also in line with a previous report that *in vivo* prolonged manipulations of RNase H and R-loops levels induce significant changes in transcriptomic landscape, and that, unexpectedly, either R-loops persistence or resolution may trigger similar responses **(Hartono *et al.*, 2018)**.

Regardless of the exact reason behind the depletion of H3K9me2, it's still hard to confirm whether this histone mark is really decreased or whether this is a result of nucleosome depletion (decrease in nucleosome occupancy). Accumulation of R-loops has been found to decrease nucleosome occupancy and enhance chromatin de-condensation **(Powell *et al.*, 2013)** while destabilization of R-loops leads to chromatin compaction **(Boque-Sastre *et al.*, 2015)**. Therefore, enrichment of H3K9me2 should be studied relative to histone H3, as it's hard to imagine a nucleosome and an R-loop to coexist at the same locus, i.e., R-loop accumulation may deplete nucleosomes in general. Anyway, the above phenotype doesn't seem to be a result of local nucleosome depletion, as it is produced in two completely contrasting conditions. In addition, based on my result in fission yeast, R-loops don't form over the affected or H3K9me2-depleted regions.

Another possible explanation is that RNase H manipulation or R-loops levels changes may affect the transcription of specific ncRNA, which may affect heterochromatin formation. It has been found that the exosome, the degradation machinery for aberrant RNA, contributes to heterochromatin assembly at centromeres and other genomic regions through ncRNA **(Reyes-Turcu *et al.*, 2011)**. Specific RNAs, including meiotic mRNAs, have been found to be targeted by RNA quality controls and elimination machinery to get involved in heterochromatin nucleation **(K. Zhang *et al.*, 2011; Zofall *et al.*, 2012)**. Specifically, preventing transcription of meiotic mRNA or depletion of RNA elimination factors, disrupts heterochromatic islands **(Zofall *et al.*, 2012)**.

Consistent with these findings, RNA elimination factors have been found to contribute to gene silencing and heterochromatin formation. A distinguished example is the MTREC complex which targets the CUTs and meiotic RNA for degradation by the exosome in fission yeast **(Shichino *et al.*, 2020; Zhou *et al.*, 2015)**. This complex is required for facultative

heterochromatin formation along with other complexes such as CCR4-NOT complex **(Cotobal *et al.*, 2015; Sugiyama *et al.*, 2016)** and Pir2 protein **(Thillainadesan *et al.*, 2020)**.

In fact, multiple pathways lead to heterochromatin formation. These include, in addition to the exosome and RNA elimination machinery, premature transcription termination **(Chalamcharla *et al.*, 2015)** and selective non-coding RNA termination at pericentromeric repeats **(Touat-Todeschini *et al.*, 2017; Vo *et al.*, 2019)**. The fact is, beside the RNAi dependent pathway, multiple RNAi-independent pathways have been found to trigger heterochromatin formation for *S. pombe* for example. An important protein involved in this pathway is the RNA-binding protein Seb1, the ortholog for Nrd1 in S. cerevisiae, which binds ncRNA at the pericentromeric repeats to recruit the histone deacetylase and repressor SHREC complex which promotes H3K9me at these regions **(Marina *et al.*, 2013)**. Seb1 protein was found to also induce long-term Pol II pausing as a signal required for heterochromatin assembly over pericentromeric repeats and triggering ectopic heterochromatin formation **(Parsa *et al.*, 2018)**. This explains why neither ago1 or dcr1 single deletion can abrogate H3K9me2 signal over pericentromeric heterochromatin.

Fission yeast, for example, has a self-amplifying and spreading heterochromatin system with many pathways for its formation **(Allshire & Ekwall, 2015; K. Zhang *et al.*, 2008)**. In addition, fission yeast has the active demethylase Epe1 which can actively remove H3K9me2 mark **(Allshire & Madhani, 2018; Zofall *et al.*, 2012)**. It should be considered that these factors are supposed to mask the effect of either R-loops accumulation or depletion.

### 5.4.5 H3K9me2 histone mark is hard to detect in fission yeast

Here I used fission yeast, one of the most powerful models for studying heterochromatin, RNAi and gene silencing **(Allshire & Ekwall, 2015; Martienssen & Moazed, 2015)**, in order to study the impact of R-loop level fluctuation (R-loop formation resolution events) on heterochromatin, which was suggested by previous reports as mentioned above. In fission yeast, heterochromatin is usually studied by detecting H3K9me2 histone modification, the hallmark of heterochromatin and most widely used heterochromatin marker for such studies in this model organism **(Cam & Whitehall, 2016; K. Zhang *et al.*, 2008; Zofall *et al.*, 2012)**. Unfortunately, I found it hard to detect total protein signal of H3K9me2 using Western blot (Chapter 3, Figure 3.3) which is consistent with what seems to be well-known in the field **(Cam & Whitehall, 2016)**. For this reason, the enrichment of this mark at heterochromatin is studied

using ChIP which produces low resolution results though. To study this histome mark genome-wide at high resolution, I used the high-resolution ChIP-exo, but I found it challenging to map H3K9me2 using original version of this technique (Chapter 3) which may represent a drawback for conducting such studies in fission yeast and suggests the need for using another marker for such studies. Apparently, this detectability issue is a common problem for all heterochromatin factors including Clr4, the key histone methyltransferase, and Swi6, the homolog of heterochromatin protein 1 (HP1) in fission yeast **(Cam & Whitehall, 2016)**. The main reason for the difficult detectability, despite their abundance over extensive regions, is that these factors exist mainly over specific genomic regions representing a small fraction of fission yeast genome **(Cam & Whitehall, 2016; K. Zhang *et al.*, 2008; Zofall *et al.*, 2012)**. This is exacerbated by the small size and single copy nature of yeast strains used. This problem presents a limitation for using fission yeast despite being a vital model organism for studying epigenetics, heterochromatin and gene silencing. Since it seems hard to use an alternative marker with fission yeast, the best solution would be to use another higher-eukaryote model organism such as *Caenorhabditis elegans*, *Drosophila*, or *Arabidopsis* which have bigger genome and bigger heterochromatin fraction. Remarkably, I found mammalian cells to be a great model to be used for such purpose. I easily detected H3K9me2 signal with Western blot and conducted a genome-wide mapping using ChIP-exo-TT. I also could detect changes in this mark between different cell types. Notably, in mammalian cells and other higher eukaryotes there are other histone marks such as H3K27me3 that may be convenient to use as a marker.

## 5.5 Conclusion

My observations form the three R-loop mapping methods don't support a link between R-loops and termination or between R-loops and heterochromatin formation in fission yeast. These methods didn't detect R-loops neither over terminators nor over heterochromatin regions. Instead, they heavily enriched R-loops over tRNA and rRNA genes embedded in heterochromatin. These R-loops didn't extend outside these genes to the surrounding heterochromatin, providing a striking evidence that R-loops are marks of active transcription but not heterochromatin formation.

Here I show remarkable depletion for H3K9me2 at heterochromatic domains in fission yeast and, also, global depletion of this histone mark in mammalian cells with RNase H depletion or overexpression, respectively. This strongly suggests that RNase H manipulation

induces global changes on transcriptome and proteome of cells. Despite dependence of this change on RNase H (at least in fission yeast), this change is independent from localized effect of R-loops at these regions.

Beside the fact that R-loop mapping need to be conducted by a different S9.6 independent method to verify current results regarding absence of R-loops over terminators and heterochromatin in fission yeast, it's important to note that current results may represent fission yeast only. Particularly, absence of R-loops over terminators can't be depended for all systems knowing that previous studies using different methods confirmed formation of R-loops over terminators and showed a relevant role in termination which is still under debate though. It would be relevant to try my methods for other biological systems especially mammalian cells to compare results obtained from mammalian cells to those from fission yeast. It's possible that multiple R-loop features and roles are not conserved among different systems. It's also critical to check results for different cell types and genotypes (WT against mutants) to check relevance of findings under physiological and genetic/pathological conditions.

My results show that different observations based on RNase H manipulation may have been inappropriately explained by others as a direct role or localized effect for R-loops. My results show that it should be considered that RNase H manipulation may induce a global effect on transcription and protein expression independent form R-loops.

# Chapter 6. Final Discussion

In this study, I sought to interrogate a previously suggested role for R-loops in dsRNA generation, transcription termination and heterochromatin formation, but on a genome-wide scale. Perturbing R-loop levels was shown to interrupt transcription termination at few terminator regions in mammalians cells **(Skourti-Stathaki *et al.*, 2011)**. Besides, R-loops have been suggested to induce RNAi-mediated heterochromatin formation through promoting antisense transcription and dsRNA formation **(Skourti-Stathaki *et al.*, 2014)** at these regions. dsRNAs are important structures required for triggering RNAi-mediated gene silencing, that's why it's important to understand the mechanism by which R-loops may induce dsRNA formation. Here, I proposed two possible models for R-loop-dependent dsRNA formation, the first is through bidirectional hybrid formation in a structure that I call double R-loops while the second is through an RNA-DNA hybrid overlapping with free asRNA. Both models assume that asRNA can be transcribed from free ssDNA of R-loop, which gains support from previous findings suggesting that ssDNA can act as a promoter for Pol II transcription **(Pan & Greenblatt, 1994; Parvin & Sharp, 1993)** and can promote RNA-DNA hybrid formation *in vitro* (Fischer lab, unpublished data). Recently, RNA-DNA hybrids have been shown to form over ssDNA of DSB ends *in vivo* **(Ohle *et al.*, 2016)**. Moreover, R-loops have been shown to induce antisense transcription over mammalian genome **(Tan-Wong *et al.*, 2019)**. Importantly, R-loop mapping data from different methods such as ChIP-exo data from fission yeast **(Ohle *et al.*, 2016)** showed double signals suggesting double R-loop formation. These multiple findings provided a launch pad for my current investigations using fission yeast.

Because of limitations for current R-loop mapping methods, a high-resolution and directional R-loop mapping was required in order to investigate both models in general and double R-loop formation in particular. Here, I used three methods which depend on the use of S9.6 antibody for immunoprecipitating R-loops from sonicated chromatin preparation. Two of these methods depended on the use of the high resolution and strand specific ChIP-exo technique with DNA-based library preparation workflow while the third depended on RNA-based library preparation workflow. For the first DNA-based ChIP-exo method, I harnessed the original ChIP-exo protocol which captures dsDNA through ligation to double-stranded DNA adaptors, so I called this method as dsChIP-exo. In order to enhance directionality, I introduced a modification for this method to capture single stranded DNA through poly-dA tailing of DNA template in a method that I called ssChIP-exo. For the third method I used

SMART technology for capturing the RNA strand of the hybrid through poly-A tailing for RNA template in a method that I called SMART-DRIPc. To enhance resolution and avoid dsRNA contamination in SMART-DRIPc, I used P1 nuclease and RNase III to digest free ssRNA and dsRNA. Data of my methods exhibited lower background, higher resolution and stronger overlap among their signals compared to external data. Signals mapped by my methods exhibited a great specificity and overlapped with sites of R-loop formation confirming the reliability of the methods and the reproducibility of the data. My methods, especially the ssChIP-exo and SMART-DRIPc led to multiple technically and biologically related findings.

***About technical findings:*** dsChIP-exo generated correlated double signals while ssChIP-exo and SMART-DRIPc exhibited unidirectional signals genome wide suggesting that dsChIP-exo isn't compatible for directional R-loop mapping. This matches previous reports that dsDNA-capturing methods such as S1-DRIP can't produce directional signals even after digesting the free ssDNA **(Wahba *et al.*, 2016)**, and only ssDNA-capturing methods such as ssDRIP produce directional R-loop signals **(Xu *et al.*, 2017)**. Findings from ssChIP-exo and SMART-DRIPc confirmed that the free ssDNA of R-loops is removed during sonication and washing steps as previously suggested **(Crossley *et al.*, 2020; Xu *et al.*, 2017)**. Regardless of directionality, my methods showed strong consistency in their results which was confirmed by spatial overlap of detected signal peaks. My methods almost share the same steps but differ in library preparation method, i.e., they share lysis, sonication, the use of chromatin input and the use of S9.6 antibody, but differ in the DNA sequence-capturing step for sequencing (adaptor ligation for dsChIP-exo or dA tailing for ssChIP-exo). The use of similar experimental conditions, especially the use of chromatin preparation seems to have limited the variation in results. A strong evidence for consistency of methods and reproducibility of results is that both dsChIP-exo and ssChIP-exo showed that fixation strengthens cumulative PCR signal for WT while weakens it for RNase H depleted cells. I found that fixation may induce variation in mapped signals, but these variations didn't change the final conclusion regarding directional R-loop formation or other results, which is consistent with previous reports that fixation doesn't affect final results **(Halasz *et al.*, 2017)**. Different from a previous study which showed that the RNA-based DRIPc may capture double signals representing dsRNA which necessitates prior treatment with dsRNA-digesting enzymes **(Hartono *et al.*, 2018)**, SMART-DRIPc generated directional signals without the enzymatic treatments which seem to have negative impact on signals. At the technical level, the difference in my results compared to other studies on the same organism can be explained by the use of sonication and chromatin preparation

(more *in situ* conditions) instead of enzymatic fragmentation for nucleic acid (completely *in vitro* conditions) **(Crossley *et al.*, 2019; K. Wang *et al.*, 2021)**. Based on my findings, I recommend the use of ssDNA-based capturing methods, chromatin input, sonication, and the omission of fixation or considering it carefully, especially for WT cells.

***About proposed models:*** Signals mapped by ssChIP-exo and SMART-DRIPc confirm co-transcriptionally directional R-loop formation as previously established using directional R-loop mapping methods for yeast, *Arabidopsis* and mammalian cells **(L. Chen *et al.*, 2017; Crossley *et al.*, 2020; Hartono *et al.*, 2018; Sanz *et al.*, 2016; Tan-Wong *et al.*, 2019; Xu *et al.*, 2017)**. Similar to these studies, multiple observations didn't support genome-wide double R-loop formation. The first observation is the absence of correlation between forward- and reverse-strand metagenic signals mapped by ssChIP-exo and SMART-DRIPc, especially for WT cells. The second is the detection of unidirectional signals genome wide. To some extent, few regions such as rDNA, tDNA repeats and mitochondrial DNA showed double R-loop signals in WT. Similarly, ssDNA-based R-ChIP detected double R-loop signals over tRNA and rRNA genes in mammalian cells **(L. Chen *et al.*, 2017)**. The third observation is the sensitivity of most of R-loop signals for RNase H in WT cells which confirms that at physiological conditions, most of R-loops are not stable and can't support dsRNA formation. One the other hand, some R-loop signals were partially or completely insensitive to RNase H in WT cells. These signals overlapped with sites of sense anti-sense transcription and coincided with dsRNA signals captured by an external DRIPc data **(Hartono *et al.*, 2018)**. These insensitive R-loops formed only on one direction, which suggests that the second model is more plausible for R-loop-dependent dsRNA formation. A paradox in this observation is that RNA-seq and NET-seq data **(Wery, Gautier, Descrimes, Yoda, Vennin-Rendos, *et al.*, 2018)** showed unidirectional transcription at some of dsRNA formation loci. It's not clear how these dsRNAs are generated or detected. Does this observation suggest dsRNA formation *in trans*?

dsRNA formation has been shown to be a feature of the genome of different organisms. In fission yeast, dsRNAs have been shown to form over 35% of pombe genome **(Hartono *et al.*, 2018)**. However, it's not clear how these dsRNAs can form, especially considering the global anticorrelation between sense and antisense transcription at nascent and mature RNA level **(Atkinson *et al.*, 2018; Murray *et al.*, 2015; Wery, Gautier, Descrimes, Yoda, Vennin-Rendos, *et al.*, 2018)**. This anticorrelation represents a critical challenge for R-loop dependent dsRNA formation as well. Although mismatched dsDNA, ssDNA constructs or DSB ends can

be transcribed and also form RNA-DNA hybrids, this situation may be different for R-loops as the presence of a hybrid may negatively affect antisense transcription and block both transcription initiation and elongation *in vitro* **(Belotserkovskii & Hanawalt, 2011**, **2015; Belotserkovskii *et al.*, 2010; Belotserkovskii, Mirkin, *et al.*, 2013; Belotserkovskii, Neil, *et al.*, 2013; Belotserkovskii *et al.*, 2017; Tous & Aguilera, 2007)** and *in vivo* **(González-Aguilera *et al.*, 2008; Huertas & Aguilera, 2003)**. Consistent with this anticorrelation, R-loop formation was found to supress transcription of overlapping asRNA in bacteria over 500 chromosomal loci **(Raghunathan *et al.*, 2018)**. This explains why asRNA induced by R-loops in mammalian cells were found to form proximally and didn't strictly overlap with RNA-DNA hybrids **(Tan-Wong *et al.*, 2019)**. Double R-loop formation and dependent dsRNA generation seem to have been driven by some observations from unfaithful R-loop mapping strategies and non-directional techniques. For example, ChIP-exo data revealed double signals and suggested double R-loop formation. The fact that this method, mainly used for mapping protein-DNA interactions at high resolution, is known to provide strand information, suggested genome-wide double R-loop formation. Also, external DRIPc mapping generated genome-wide double signals which were found to represent exosome sensitive dsRNA and could be eliminated by RNase III but not RNase H **(Hartono *et al.*, 2018)**. Importantly, my SMART-DRIPc didn't detect such signals.

***About response of strand signals to RNase H deletion:*** A remarkable finding is that forward strand signals were more sensitive to RNase H in WT cells (they are depleted in WT and strongly increase with RNase H deletion) compared to reverse strand signals, i.e., reverse stand signals exhibited insensitivity toward RNase H and got slightly stronger in WT compared to RNase H-depleted cells. Importantly, the opposing change in signal strengths over both strands match the known anticorrelation for sense-antisense transcription. Relevant to this finding, RNase H deletion decreased R-loop levels over reverse strand (template strand) of sense-transcribed translation elongation and ribosome biogenesis genes. This implies a decrease in transcription rate of these genes which suggests also a decrease in their expression rate and, accordingly, an impact on cellular protein synthesis processes. This effect doesn't seem to depend on localized enzymatic activity of RNase H which means it happens independent from local R-loops, but due to a global effect for RNase H deletion (or global R-loop accumulation) which seems to affect transcription profile of these cells. For better understanding for the cause-effect correlation, transcriptome profiling is required to study

potential changes in transcription upon RNase H deletion over both DNA strands and to explain changes in R-loop levels as R-loop formation is mainly a function of transcription.

***About termination and heterochromatin:*** My methods didn't enrich R-loops at gene terminators compared to gene bodies nor detect R-loops over heterochromatin in fission yeast. Instead, R-loop signals exhibited uniform distribution along gene bodies. The fact that all my methods detected strong R-loops that are strictly confined to highly-transcribed tRNA genes embedded in pericentromeric heterochromatin and similarly over rRNA genes at chromosome III telomers (a heterochromatic domain), without extending outside their gene bodies, confirms that R-loops are marks of active transcription rather gene silencing. My results strongly agree with results of a recent study from Chedin lab which found that R-loops don't form over terminators, but strongly correlate with marks of active transcription in fission yeast **(Hartono *et al.*, 2018)**. Importantly, the difference in observations regarding R-loop formation over terminators and heterochromatin in fission yeast can be explained by difference in resolution of methods used. Similar to DRIPc, all my methods have high resolution compared to previous DRIP and ChIP methods which detected R-loop over these regions **(Bronner *et al.*, 2017; Castel *et al.*, 2014; Nakama *et al.*, 2012)**. For mammalian cells, Chedin lab detected R-loops at terminator regions, but found that R-loops correlate with transcription activation marks rather than H3K9me2 or other heterochromatic marks **(Sanz *et al.*, 2016)**. In this regard, while Proudfoot lab suggested a role of R-loops in transcription termination and heterochromatin formation in mammalian cells, they couldn't confirm enrichment of R-loops over terminators or detect heterochromatin in mammalian cells using RR-ChIP method **(Tan-Wong *et al.*, 2019)**. The fact that none of my methods enriched R-loops over gene terminators or detected R-loop over heterochromatin in fission yeast, didn't encourage further functional examinations to study the role of R-loops in transcription termination or heterochromatin formation.

***About response of heterochromatin to RNase H manipulation:*** Strikingly, I found that perturbing R-loop levels through RNase H deletion or overexpression, disrupts H3K9me2 signal at heterochromatin in fission yeast and globally in mammalian cells, respectively. It's hard to explain these results knowing that R-loops don't form over heterochromatin in fission yeast. This can't be explained by a direct impact for localized R-loop level change or as consequence for local enzymatic activity of RNase H. Yet, this would be rather explained by an impact for global R-loop level changes or as a result of RNase H level imbalance which

may have led to multiple biological deficiencies in transcription and protein synthesis processes.

***Based on findings from my study, I find it important to stress that some technically and biologically related facts in the R-loop field have to be critically considered.*** As a rule, R-loop mapping results should be explained by methodology and experimental conditions and, importantly, biological system, model organism and genotype as many of R-loop features and functions may not be preserved under physiological conditions or in all organisms. For example, my results obtained for fission yeast can't be generalized for all systems and *vice versa*. Examples for variable features among different organisms are link to GC or AT content, correlation with open chromatin and chromatin marks, association with heterochromatin, and enrichment over specific genic compartments, especially promoters and terminator regions.

Apparently, the enrichment of R-loops at terminator regions and the involvement of R-loops in termination are among the highly debated areas of R-loop research. My results from fission yeast didn't support a relevant role for R-loops. However, the fact that R-loops have been found to form over terminator regions in mammalian cell and also to play a role in termination (with some reservations) suggest that further investigation is required to confirm these findings for mammalian cells against yeast using my current methods. This is important for a comparative and more relevant conclusion. Also, the fact that different methods may produce discrepancies in results suggests that my results for fission yeast should be checked using experimental approach with more native *in situ*/*in vivo* conditions such as RNase-H based methods. For these reasons, new tools should be developed to match the need for more precise mapping and better functional analysis. The use of multiple methods with different capturing probes (either S9.6 or RNase H) should be pursued, but this may show a gap between results.

It's critical to carefully interpret results based on response toward cellular RNase H manipulation (deletion or overexpression). RNase H manipulation may exhibit an effect that's not directly dependent on R-loop enrichment over a specific locus, i.e., it may not reflect a localized enzymatic activity (R-loop degradation effect) for RNase H, but may be due to either a global effect for R-loop level fluctuation or RNase H deficiency (level imbalance). The fact that RNase H depletion was found to affect expression of protein synthesis and ribosome biogenesis which was confirmed by a previous study and can be inferred from my R-loop mapping data, suggest that RNase H depletion may affect transcriptome and proteome of cells

independent from R-loops. Consistent with this, RNase H depletion or overexpression seems to disrupt heterochromatin where R-loop signals weren't detected. These results make it irrelevant to make a conclusion about a direct role of R-loops in termination and heterochromatin formation.

Obviously, existence of double R-loops can't be confirmed by correlation analysis. Even if double signals are detected and a strong correlation is revealed between both signals, a more dedicated approach is required to show that these signals form over the same allele and in the same cell. Since current mapping methods capture signals for a population of cells, these signals may be captured for different loci or from different cells (one cell forms a forward while another forms a reverse signal). The most confirmative approach I can see is R-loop mapping for a single cell which may be very challenging as it requires very sensitive methods. For this, I suggest the use of either SMART-seq or CUT&Tag approach in combination with S9.6 or RNase H-based R-loop capturing strategy. As an alternative for single-cell approaches, I can suggest the use of two methods in parallel, which may be insightful, but not definitive. For example, it's possible to use a method that can detect hybrids (such as S9.6- or RNase H-based R-loop mapping methods) along with another method that can detect free ssDNA (such as bisDRIP or replication protein A (RPA)-based methods) for detecting different signals. For unidirectional R-loops, these methods should show opposite signals while for double R-loops, only S9.6- or RNase H-based, but not RPA- or bisDRIP-based methods should produce a singal. This approach may require stringent normalization and very complicated analyses. This may provide better information compared to the use of multiple, but similar methods (methods which only detect hybrids or those which only detect free ssDNA).

Importantly, the competency of mapping methods and library preparation strategies for capturing and sequencing double R-loops need to be tested, and possibility of inhibition by any conditions need to be excluded. Similarly, the directionality of used methods has to be validated. This can be practically achieved using artificial control and spike in R-loops. The use of artificial R-loop controls of known sequences and different structures (lengths, overhang, and different sequence composition) may be helpful for precise R-loop mapping. Artificial R-loops may provide a clear evidence how different methodologies work by comparing the final results to predicted outcome. Three-stranded artificial R-loops are more relevant compared to RNA-DNA hybrids and can be used for many purposes; for normalization and quantitative analysis; for confirming resolution and directionality of methods; for testing

bias in library preparation strategies; for revealing the effect of lysis, fragmentation and sonication; for studying stability of R-loops against different enzymes; and for studying transcriptiton over ssDNA in R-loop *in vivo*. Especially, double R-loops are required to check whether current methodologies can capture both RNA-DNA hybrids on both directions. Because, preparation of stable three stranded R-loops is challenging, here I introduced the use of bubble dsDNA with middle non-complementary sequence as a method for preparing permanently stable R-loops without special nucleotides, formamide, high temperature or modifying agents. This strategy is insightful for preparing stable double R-loops.

## 6.1 Prospective

First of all, R-loop mapping needs to be conducted using another directional method, but with more *in vivo* conditions (such as RNase H-based R-ChIP or nuclease targeted CUT&Tag) for fission yeast to check reproducibility of results obtained by my methods. Also, R-loop mapping should be done for mammalian cells using my methods to compare results obtained for fission yeast, especially R-loop formation over terminators and heterochromatin. To check changes in nascent transcript level and study termination in respect to R-loop level changes, NET-seq can be used, but again results should be carefully interpreted (as a direct effect for R-loop or a response to RNase H manipulation), especially if R-loops weren't detected over these reegions. It's important to map dsRNA using a dedicated methodology and a specific antibody such as J2. siRNA mapping can be done to study the effect of R-loop level perturbation on siRNA and, accordingly, RNAi pathway. This may be another way to study impact of R-loops on gene silencing independent from heterochromatin. Nascent and mature transcriptome profilings (NET-seq and mRNA-seq) are required to understand the molecular consequences of RNase H deletion, especially the increase and decrease of signals, respectively on forward and reverse strands. This is critical to conclude whether this effect is due to decrease of sense (increase of antisense) transcription or due to increase of R-loop levels over forward strand which antagonistically affect reverse strand signal. To confirm double R-loop formation over specific regions, single-cell R-loop mapping should be done for a definitive evidence. SMART-seq or CUT&Tag are the best candidates for library preparation for their sensitivity. They can be used with either RNase H- or S9.6-based capturing method. Single-cell mapping can be done quantitaivley using control R-loops as spike-ins and strandards for normalization.

# Appendix. A Novel Artificial Phenotype Uncoupling mRNA and Protein Expression

## A.1 Introduction

### A.1.1 Interplay between different factors for control of gene expression

Beyond packaging and protecting the genome, the chromatin plays an essential role in regulating different DNA-related nuclear processes, through limiting and regulating access of different molecules to the underlying DNA sequence **(B. Li *et al.*, 2007)**. The nucleosome, the basic building unit of chromatin, consists of four heterodimers of histone proteins H2A, H2B, H3 and H4 forming a basic octameric core. Around this core is wrapped nearly 147 base pairs of double stranded DNA helix contacting positively charged histone residues at about 14 points. These multiple contact points make the nucleosome-DNA one of the most stable complexes known among cellular protein-DNA complexes **(B. Li *et al.*, 2007)**.

Despite the strong nucleosome-DNA binding, the nucleosome has been found to be a very dynamic unit that is affected by multiple processes and continuously modulated by different factors that change its identity and association with DNA, and impact local chromatin state. Three main features define the state of a nucleosome and characterize the chromatin at a specific genomic locus as either open and accessible, or blocked and inaccessible local chromatin. These features are: (1) nucleosome positioning which describes the exact position of a nucleosome over a genomic locus; (2) nucleosome occupancy which refers to the frequency that a specific position is occupied by a nucleosome within the cell population; and (3) histone exchange, or nucleosome turnover which indicates the rate of exchange between chromatin-bound and soluble histones or, more widely, the rate of disassembly and assembly of nucleosomes. These features characterize the local chromatin landscape over the genome **(Hennig & Fischer, 2013)**.

Work in the field has revealed a cross talk between multiple factors for regulating chromatin dynamics and gene expression. For instance, histone modifying enzymes and other epigenetic factors such as chromatin remodelers and histone chaperones cooperate for

modulating the above nucleosome features, especially, the assembly and disassembly of nucleosomes (nucleosome turnover). On the other hand, increased nucleosome turnover and histone exchange rate can alter the epigenetic state of chromatin by replacing modified histones or incorporating different histone variants with different physical and chemical properties **(Venkatesh & Workman, 2015)**. The DNA sequence itself participates in this crosstalk as the consensus DNA sequence can permit the binding of specific proteins and transcription factors. Moreover, specific DNA sequences alter the DNA-binding affinity of nucleosomes and affect nucleosome stability/occupancy **(Murr, 2010; Venkatesh & Workman, 2015)**. The ultimate goal of these processes is to control the accessibility of DNA and to regulate different nuclear processes such as DNA replication, transcription, recombination and DNA damage repair.

### A.1.2 Chromatin represses pervasive transcription and maintains genomic stability

The basic chromatin architecture over transcription units, has been found to be conserved among eukaryotes. At the promoter region, the chromatin exists in an open state with a nucleosome depleted region (NDR) surrounded by two precisely positioned nucleosomes of a high turnover rate. Nucleosomes at this region have highly acetylated histone residues while histone H3 lysine 4 is mostly trimethylated as a mark of active promoters **(Hennig & Fischer, 2013; Rando & Chang, 2009; Venkatesh & Workman, 2015)**.

Over gene bodies, nucleosomes have a relatively lower acetylation level and lower turnover rate, compared to promoter region, and form a compact array with roughly uniform spacing. Different lysine residues such as histone H3 lysine 27 in mammals exist in a methylated form. Similarly, histone H3 lysine 36 is methylated and the level of methylation correlates with the level of transcription **(Hennig & Fischer, 2013; Rando & Chang, 2009; Touat-Todeschini *et al.*, 2012; Venkatesh & Workman, 2015)**.

The abovementioned factors and chromatin features play a crucial role in demarcating functional parts and transcriptional units within the genome, allowing transcription initiation only at canonical promoters, and preventing unwanted transcription events over gene coding regions. Perturbing nucleosome spacing, positioning, occupancy or turnover rate, or affecting acetylation or methylation level of nucleosomes over transcription units, can lead to transcription initiation at points outside canonical promoter regions. This phenomenon is described as “pervasive or cryptic transcription”, while the newly activated non-canonical promoters are called cryptic promoters. The resulting non-coding transcripts are called cryptic

unstable transcripts (CUTs). Chromatin structure along with RNA degradation machinery effectively control and regulate the level of cryptic transcripts **(Hennig & Fischer, 2013; McDaniel & Strahl, 2017; Touat-Todeschini *et al.*, 2012)**.

Pervasive transcription may negatively impact cellular homeostasis. At the transcriptional level, it may induce formation of R-loops of deleterious impact on genomic integrity. This has been proven for some exosome-sensitive cryptic transcripts **(Pefanis *et al.*, 2015)**. Intriguingly, at the post-transcriptional and translational level, many cryptic and non-coding transcripts have been found to be translationally competent **(Atkinson *et al.*, 2018; Cheung *et al.*, 2008; Duncan & Mata, 2014; Huraiova *et al.*, 2020)**. These non-coding transcripts may contain ORFs that, when expressed, can produce truncated or alternative protein isoforms **(Wei *et al.*, 2019)**. These non-wild type proteins may exhibit different functionality or increase the risk for tumorigenesis.

### A.1.3 Chromatin factors repressing pervasive transcription

Several chromatin factors and protein complexes have been found to repress pervasive transcription **(Hennig & Fischer, 2013; McDaniel & Strahl, 2017; Touat-Todeschini *et al.*, 2012)**. However, the mechanisms underlying transcriptional repression by these factors aren't clearly understood. Besides, the interactions between them are still areas of debate. These complexes, depicted in Figure A.1, are:

#### *A.1.3.1 Chd1 chromatin remodeling complex*

Chd1 is a chromatin remodeling complex that regulates nucleosome spacing and positioning to maintain regular nucleosome array over gene coding regions. Deletion of hrp1 and hrp3, Chd1 chromatin remodelers, perturbs the uniform nucleosome spacing and positioning, which leads to the appearance of nucleosome depleted regions (NDRs) over transcribed genes. These NDRs are believed to trigger cryptic promoters and increase pervasive transcription **(Hennig *et al.*, 2012; Pointner *et al.*, 2012; Shim *et al.*, 2012; Smolle *et al.*, 2012)**. Deletion of Chd1 chromatin remodelers, also, increases histone turnover over gene coding regions (**Smolle *et al.*, 2012**).

### *A.1.3.2 FACT complex and other histone chaperones*

Histone chaperones such as the FACT complex, Asf1, HIRA complex and Spt6 along with other transcription elongation factors have an important role in maintaining nucleosome occupancy behind the wake of RNA Pol II **(Belotserkovskaya *et al.*, 2003; Formosa, 2008; Formosa *et al.*, 2002)**. FACT mutants are characterized by decreased nucleosome occupancy (nucleosome depletion) and increased nucleosome turnover rate over transcribed loci. However, the regular nucleosome array and uniform spacing aren't affected (**Hennig *et al.*, 2012**).

### *A.1.3.3 Rpd3s HDAC complex*

The Rpd3S complex in budding yeast, or Clr6 Complex II in fission yeast, is a histone deacetylase complex (HDAC) that removes acetylation marks from different histone residues **(Carrozza *et al.*, 2005; Keogh *et al.*, 2005; Krogan *et al.*, 2003; Nakayama *et al.*, 2003; Nicolas *et al.*, 2007)**. Rpd3S/Clr6-CII HDAC mutants are characterized by increased acetylation level (hyperacetylation), high histone exchange over gene coding regions, and severe accumulation of cryptic transcripts **(Hennig *et al.*, 2012; Nakayama *et al.*, 2003)**.

### *A.1.3.4 Set2 histone methyl transferase*

Set2 is a histone methyl transferase (HMT) that is responsible for different levels of methylations (mono, di or tri) for histone H3 lysine 36 (H3K36) residue. Set2 has been found to be recruited by the elongating RNA Pol II, while H3K36me has been identified as a mark of active transcription. Moreover, the level of methylation has been found to correlate with the level of transcription **(Krogan *et al.*, 2003; Suzuki *et al.*, 2016; Xiao *et al.*, 2003)**. In a recent study, a strong terminator was inserted in the first third of a long gene of ~10 kb and the methylation level was evaluated at regions before and after the terminator. H3K36 trimethylation was detected only at the transcribed part but didn't spread beyond the terminator **(Feil *et al.*, 2015)**. Set2 deletion activates cryptic promoters and increases pervasive transcription **(Suzuki *et al.*, 2016; Venkatesh *et al.*, 2016)**.

These findings suggest that deficiency of different individual factors can lead to increase in pervasive transcription levels. This may occur through revealing cryptic promoters and unleashing opportunistic transcription over the underlying DNA sequence. However, the minimal requirements for cryptic promoter formation and activation are still unknown. Apparently, the underlying DNA sequence itself may contribute to their formation. Importantly, cryptic promoters may differ in their reliance on specific chromatin factors for their regulation, i.e., their responses toward different molecular cues are different **(Pattenden *et al.*, 2010; Wei *et al.*, 2019)**. Precisely-positioned nucleosomes, high nucleosome occupancy and low-rate histone turnover can be thought of as guardians making a barrier against Pol II opportunistic permeation and transcription initiation. Meanwhile, Chromatin remodelers, histone chaperones and histone deacetylases can be imagined as higher order officials required for coordinating these processes to maintain balance and order and avoid unwanted harmful activities. Finally, DNA and cryptic promoters can be envisaged as hidden pervasion points of different features and regulatory rules.

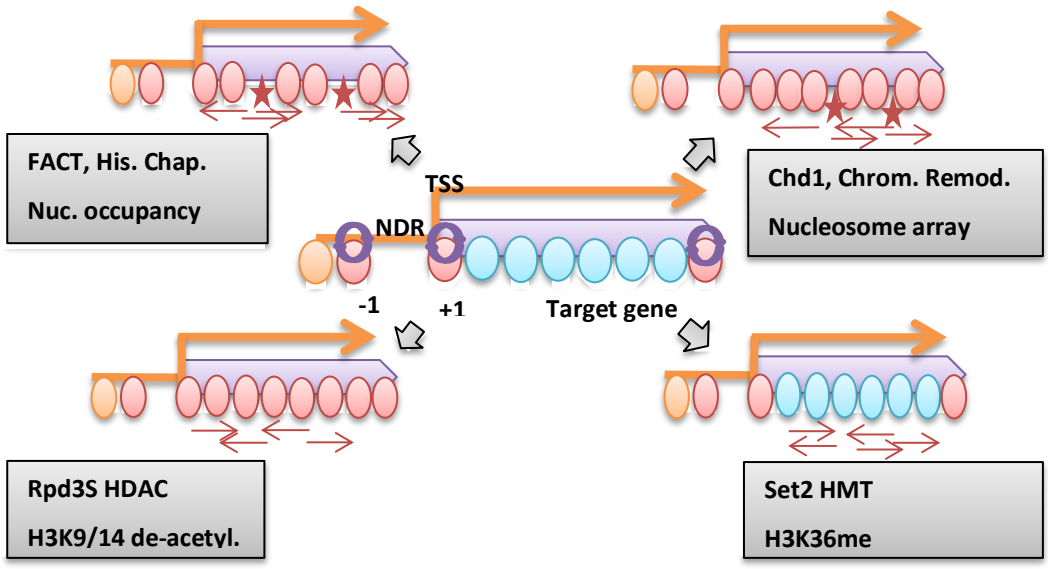



**Figure A.1. Factors modulating chromatin architecture and preventing pervasive transcription**. The middle carton depicts wild type state while the surrounding ones depict mutant features for each factor. Main function of each factor is included in a grey box. Asterisk and gaps between nucleosomes represent cryptic promoters, red arrows represent cryptic non-coding transcripts. Green and red colour codes for low turnover and high turnover nucleosomes, respectively. Inspired from **Hennig and Fischer (2013)** and **Touat-Todeschini *et al.* (2012)**.

### A.1.4 Set2-mediated H3K36 methylation and cellular functions

Interestingly, Set2-mediated H3K36me has been implicated in different molecular processes including alternative splicing, dosage compensation, X-chromosome inactivation, DNA replication and repair, exon definition, DNA methylation, and transmission of gene expression memory from parents to offspring during development **(McDaniel & Strahl, 2017; Venkatesh & Workman, 2013; Wagner & Carpenter, 2012)**. Notably, Set2 has also been linked to many diseases. Its mutations are the 4th ranked in Clear Cell Renal Carcinoma (ccRCC). Moreover, Set2 has been implicated in many cancers such as breast cancer and acute leukemia, identifying it as a potential target for onco-therapeutics **(J. Li *et al.*, 2016)**.

Set2-mediated H3K36me has been suggested to play multiple roles in chromatin dynamics (Figure A.2). However, the mechanism for H3K36me-mediated repression of pervasive transcription is quite controversial. Earlier studies attributed the repression effect to the recruitment of the Rpd3S HDAC complex (or Sin3B complex in mammals) **(Carrozza *et al.*, 2005; Keogh *et al.*, 2005)**. This explanation is currently highly debated as Rpd3S HDAC complex has been found to be directly recruited by CTD of elongating Pol II. Moreover, Rpd3S HDAC complex localization isn't affected by set2 deletion in yeast **(Drouin *et al.*, 2010; Govind *et al.*, 2010)**. Besides, it has been shown that acetylation levels are only slightly affected in set2 mutants **(Suzuki *et al.*, 2016)**.

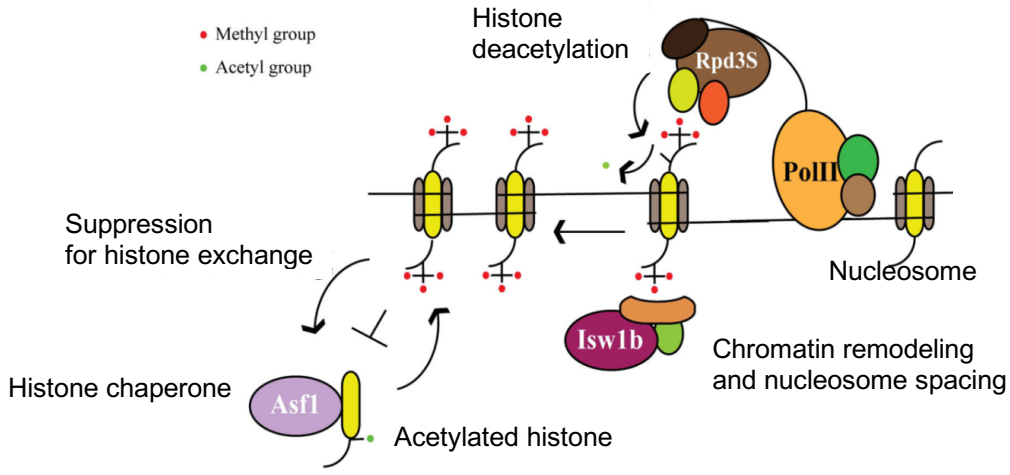



**Figure A.2. Suggested roles for Set2-mediated H3K36me in chromatin dynamics**. Set2-mediated H3K36me has been suggested to inhibit histone chaperone Asf1-mediated histone exchange, to recruit Rpd3S HDAC for histone deacetylation and to recruit chromatin remodelers required for maintaining regular nucleosomal array and uniform spacing. Adapted from **Venkatesh and Workman (2013).**

Separately, set2 deletion was found to increase the rate of nucleosome turnover at gene coding regions, suggesting that Set2-mediated H3K36me is important for maintaining low nucleosome turnover rate and repression of pervasive transcription **(Venkatesh *et al.*, 2012)**. However, unpublished results from Fischer lab showed that the nucleosome turnover rate is poorly affected in *set2Δ* mutants. Consistent with this, it has been recently reported that H3K36 methylation by Set2 and nucleosome turnover are mutually-independent factors, but are correlated with transcription **(Ferrari & Strubin, 2015)**. Despite these exciting trials to understand the correlation between H3K36 methylation and nucleosome turnover, the correlation between nucleosome turnover and pervasive transcription is not clearly understood. Increasing histone turnover rate beyond a specific threshold may increase pervasive transcription through revealing cryptic promoters and increasing DNA accessibility.

## A.2 Original hypothesis, research motivation and aims

Genuinely, I was interested in understanding the mechanism by which Set2 can repress pervasive transcription. Since Set2-mediated H3K36 methylation is known to be a mark of active transcription elongation, and, at the same time, to prevent pervasive transcription within the gene body, I hypothesized that recruiting Set2 to an active promoter will inhibit transcription initiation at this promoter. To approach this, I harnessed different targeting systems to artificially tether Set2 to different promoter types driving different reporter genes. I depended mainly on the versatile tetracycline/doxycycline controlled and inducible tet system to tether Set2 to these promoter sequences. I speculated that H3K36 methylation by Set2 would have a negative effect on transcription initiation and expected that the tet system is neutral efficient system that can be faithfully employed for this purpose.

## A.3 Results

### A.3.1 Tethering TetR-Set2 to TATA*cyc1p* abrogates the protein signal of reporter genes, but doesn't impact their RNA level

To target Set2 to a promoter-reporter system and study the speculated effect on transcription initiation and gene expression, I harnessed the tetracycline-controlled trans-suppressor (tTS) system. For this, I fused fission yeast set2 to the DNA-binding domain sequence of the tet repressor (tetR) and transfected the fusion construct in fission yeast. this construct is constitutively expressed under control of human CMVp to produce TetR-Set2 chimeric protein. Next, I transfected another construct containing the tet response elements (TRE) which consist of seven tet operon repeats (tetO7) driving minimal cyc1 TATA promoter (TATA*cyc1p*) to control the expression of the reporter genes. With normal growth conditions, in the absence of doxycycline, or tetracycline (-Dox, or -Tet), TetR binds to the tetO7 repeats of the TRE, constitutively tethering Set2 to the TATA*cyc1p* sequence (+Set2). I call this the tet-on condition as the TetR is in the active tethering mode. Adding Dox (+Dox) sequesters TetR and releases Set2 (-Set2) from the promoter, a condition which I also call tet-off as the tethering mode is turned off (Figure A.3). As a reporter gene, I first used FLAG-tagged lacZ for quick detection of its protein signal.

Surprisingly, a protein signal for LaZ wasn't detected at all in absence of Dox (-Dox/+Set2), but strikingly, a strong protein signal was detected at the expected molecular weight of LacZ for sample incubated overnight with Dox (+Dox/-Set2). This clear observation, obtained with multiple technical and biological replicates, strongly evidenced that tethering TetR-Set2 to the promoter, completely abrogated LacZ protein signal (Figure A.4B).

Incredibly, RT-qPCR analysis revealed that the RNA level of lacZ was roughly the same for both the tet-on and tet-off samples, and even tended to be higher for the tet-on (-Dox/+Set2) condition (Figure A.4C). These findings suggest that the tethered construct affects the protein expression, but not transcription, and points toward a novel phenotype that I sought to confirm and verify next. These observations also suggest a post-transcriptional mode of action for the tethered fusion protein.

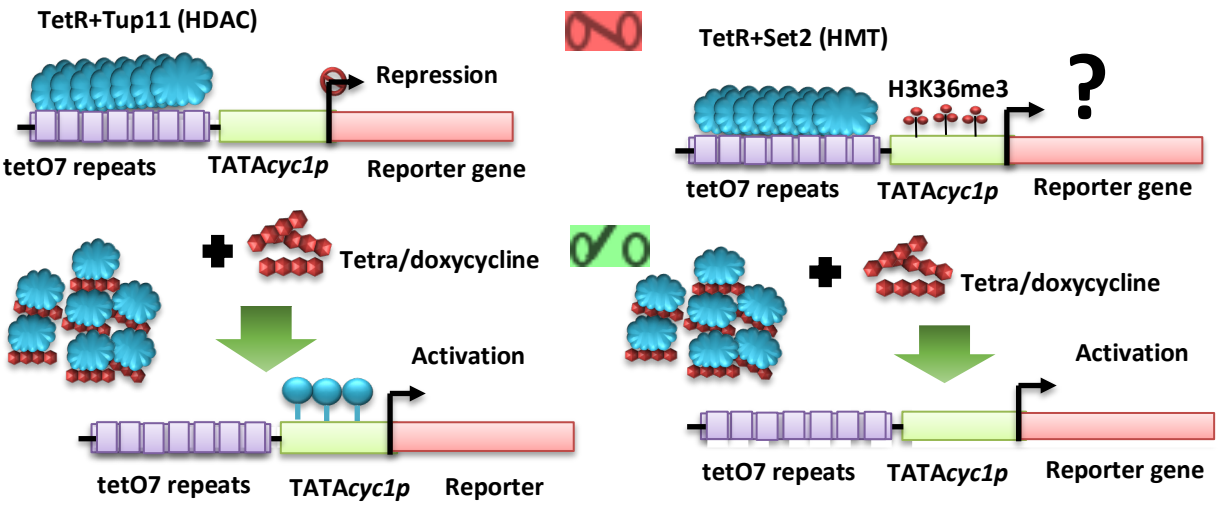



**Figure A.3. Overview of the adopted tetracycline trans-suppressor system (tet-tTS).** In the absence of dox, the TetR protein binds to the teto7 repeats upstream the TATA elements of cyc1 promoter, tethering the effector protein. Adding doxycycline, sequesters the TetR protein which gets released with the recruited proteins from the promoter region. A histone modifying enzyme, either Tup11 (left) or Set2 (right) is tethered to a specific promoter using this strategy.

### A.3.2 TetR-Set2-induced abrogation of protein signal isn't reliant on the sequence or structure of reporter genes

To confirm and validate the results produced above, I sought to study the impact of tethering TetR-Set2 to the same promoter, but with different reporter genes. Therefore, I used strains with the same system, but with either FLAG-tagged fyH3.1 (fission yeast histone H3.1 gene) or gga21 (fission yeast intronic non-essential gene) instead of lacZ. The later was used to test the effect for an intronic gene. Consistent with the above results, tethering TetR-Set2 to the TATA*cyc1p* abolished the protein signal of H3, and severely depleted the protein signal of gga21 (Figure A.4D, F). Surprisingly, the RNA level of H3 for the tet-on (-dox/+Set2) sample was lower than that of the tet-off (+dox/-Set2) sample by ~25% (Figure A.4E). Intriguingly, this slight decrease doesn't match the decrease in protein signal which completely disappeared. By other words, the absence of the protein signal can't be explained by the decrease of RNA level. My results show that this phenotype is reproducible with different reporter genes and isn't dependent on the gene sequence or structure.

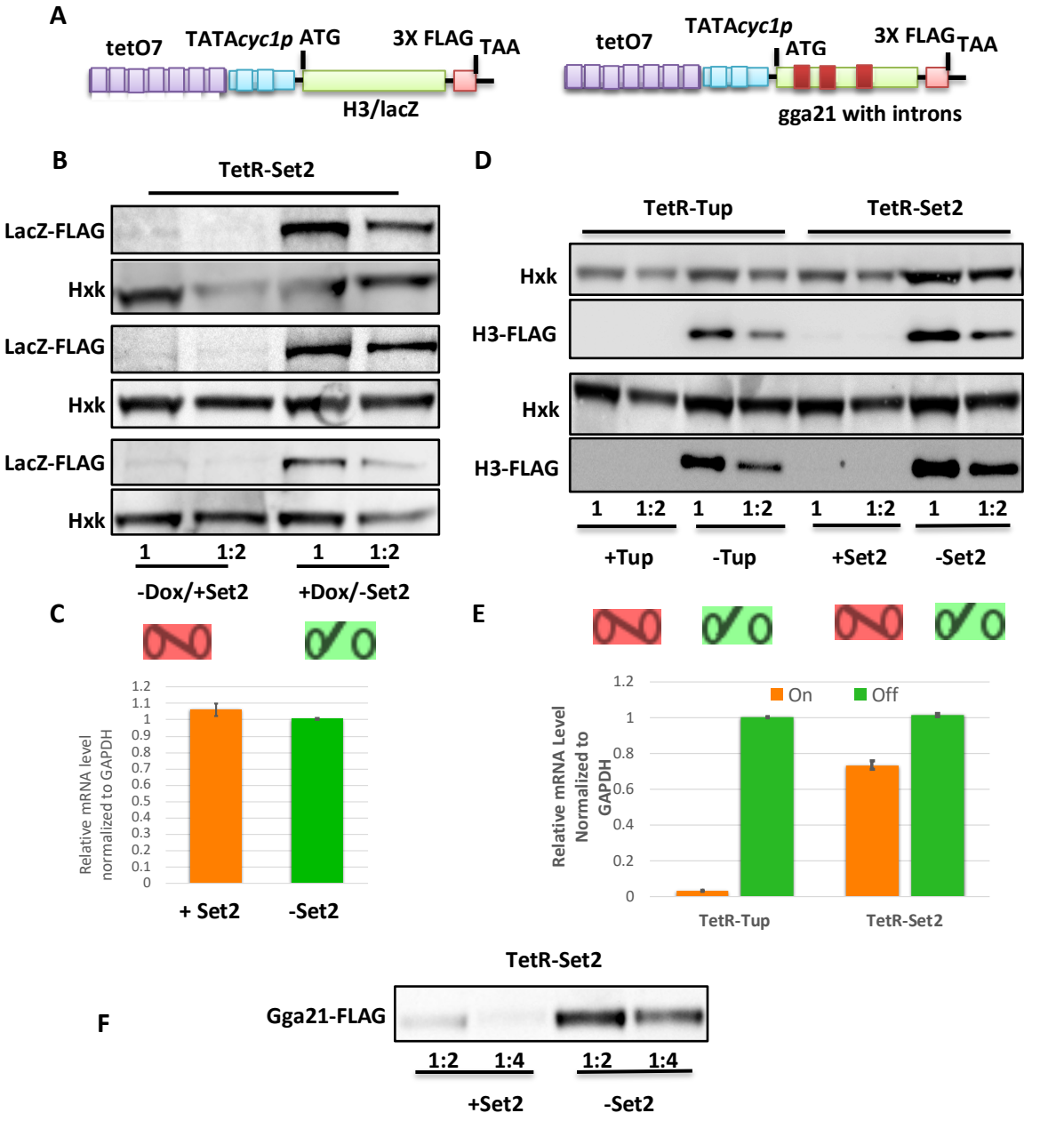



**Figure A.4. Tethering TetR-Set2 to TATA*cyc1p* abrogates the protein signals of driven reporter genes independent form their RNA level.** Western blot and qPCR analyses showing the protein signals and RNA levels of the reporter genes with and without adding 10 µM doxycycline to overnight grown cultures. Western blot and qPCR experiments were done from the same culture for each condition. Protein signal for target genes was detected using anti-FLAG antibody while Hexokinase (Hxk) was used as a control for Western blot signal. To confirm specificity of the qPCR signal, primers were designed to bind to the FLAG sequence and to a gene-specific sequence. **(A)** schematic representation for the heterologous promoter-reporter construct for either H3/lacZ (left) or intronic gga21 (right). **(B)** Western blot analysis for lacZ-FLAG protein signal showing that lacZ protein signal disappears in absence of dox in a strain carrying TetR-Set2. **(C)** qPCR analysis for lacZ-FLAG RNA level at the same conditions in **(B),** showing that the RNA level of lacZ isn't reduced. **(D)** Western blot analysis for H3-FLAG protein signal affected by the tethering or release of either TetR-Set2 or TetR-Tup**. (E)** qPCR analysis for H3-FLAG RNA level at the same conditions in **(D)**. TetR-Tup system was used as a control system with a known phenotype in order to verify the phenotype of TetR-Set2. **(F)** Western blot analysis for gga2l-FLAG protein signal affected by the tethering or release of either TetR-Set2. Error bars represent standard deviation. Different dilutions were used for the protein extract to differentiate Western blot signals.

### A.3.3 TetR-Set2-induced protein signal abrogation phenotype is different from other regular phenotypes

The tet system is a versatile and widely used system for the control of gene expression, usually, through fusing DNA-binding domain of TetR to a protein called Tup11Δ70. Tup11Δ70 is a histone deacetylase protein missing 70 amino acids representing a non-specific DNA-binding domain. Tethering this protein to a promoter region is known to completely shut down the transcription of a gene of interest which, accordingly, depletes the protein signal of target gene **(Zilio *et al.*, 2012)**. Initially, I decided to use the TetR-tup11Δ70 in combination with the H3-FLAG system, only, as a control to verify results obtained for TetR-Set2. Explicitly, tethering TetR-Tup11Δ70 to TATA*cyc1p* decreased the RNA level of H3 to the background level and eliminated the protein signal as well (Figure A.4D, E). This confirms that the phenotype conferred by TetR-Set2 is different from that of TetR-Tup∗. Based on the fact that the RNA level isn't the key for explaining the abrogation of the protein signal by TetR-Set2, I speculated that the transcription level of reporter genes isn't affected. This also excludes the possibility of recruiting a histone deacetylase by Set2 or a decrease in turnover rate at the promoter.

### A.3.4 Tethering TetR-Set2 to native TATA-less promoters doesn't affect the protein signal of their endogenous genes

Promoters are known to exhibit differences in the DNA sequence of their core elements, nucleosome features and associated histone modifications, the targeting molecular mechanisms, and, more importantly, in their response to different factors and molecular cues **(de Jonge *et al.*, 2017; Kubik *et al.*, 2015; Kubik *et al.*, 2017; Natsume-Kitatani & Mamitsuka, 2016; Watanabe & Kokubo, 2017)**. Keeping this in mind, I wanted to examine the promoter requirements for this TetR-Set2-induced protein abrogation effect. First, I wanted to examine the consequences of tethering TetR-Set2 to native homologous TATA-less promoters, to test whether these promoters will show the same response. To address this, I

---

∗ N.B. It's important to note that I used TetR-Tup only as a control with a known phenotype to which I can compare the phenotype of TetR-Set2 to prove that the mechanisms are different (TetR-Set2 effect is independent of RNA level). Later, it came to my knowledge that the TRE response elements (TATA*cyc1p*-H3) of the control system are inserted at a locus different from those of TetR-Set2. This adds another layer of difference that I carefully considered before driving a conclusion regarding the mechanism behind. Regardless, I believe the TetR-Set2 is a novel phenotype which is exciting enough to understand.

performed *in-situ* integration for tetO7 repeats, without the TATA*cyc1p*, upstream the native promoter sequence of two non-essential genes, emc1 and gga21 in a fission yeast strain that already has the TetR-Set2 fusion system (Figure A.5A). I tagged these genes with FTP tag for detecting their protein signal. I used the intronic gga21 gene to test whether tethering TetR-Set2 will affect the splicing of the pre-mRNA of this gene.

Previously it has been shown that a single Set2 molecule is enough to show promoter proximal transcription repression **(Strahl *et al.*, 2002)**. To verify this and test the copy number requirements for the tethered Set2 to produce this phenotype, I decided to use the catalytically dead Cas9 protein (dCas9) to target Set2 to the same native promoters (Figure A.5B). I fused set2 to the 5' end of dCas9 to be constitutively expressed under control of TDH3 promoter. After integration in yeast, this construct should be expressed into a chimeric protein of Set2 fused to N-terminus of dCas9. I designed specific gRNA for dCas9 to target either FTP-tagged gga21 or emc1 gene. The main difference between dCas9 and TetR of the tet system is that the former doesn't require an additional sequence to be integrated upstream the promoter sequence of target genes. Different from the tet system which can tether multiple copies of Set2, dCas9 will tether only one copy of Set2 to the target sequence. For comparison and as a control, I used dCas9 alone to represent the wild type condition.

Notably, I didn't detect a clear change in the protein signal of targeted genes, compared to the control, neither with the TetR-Set2 nor with Set2-dCas9 system (Figure A.5A, B). These findings suggest that the phenotype produced by TetR-Set2 is dependent on the promoter type of the targeted gene*. Since no effect was showed using the tet system which target multiple copies, it was not applicable to conclude whether one molecule is required to induce an effect.

Next, I used Set2-dCas9 system to test whether Set2 may affect splicing of gga21 mRNA. For this, I synthesized cDNA from total RNA then performed PCR amplification using a primer pair that flanks the third intron (binds to the end of second and start of fourth exon) of gga21. Interestingly, PCR amplification produced multiple amplicons for Set2-dCas9 but not dCas9

* It's critical to note that regardless of the promoter type, another variable may contribute to the difference in the effect of TetR-Set2 on these two promoter categories. This factor is the genomic location of the response elements. Obviously, the heterologous TATA*cyc1p*-reporter system is a transgenic system inserted at a genomic locus different from that of the native TATA-less promoter-gene.

sample (Figure A.5B). This observation implies that Set2 may have affected the splicing of this gene which may require further careful investigation.

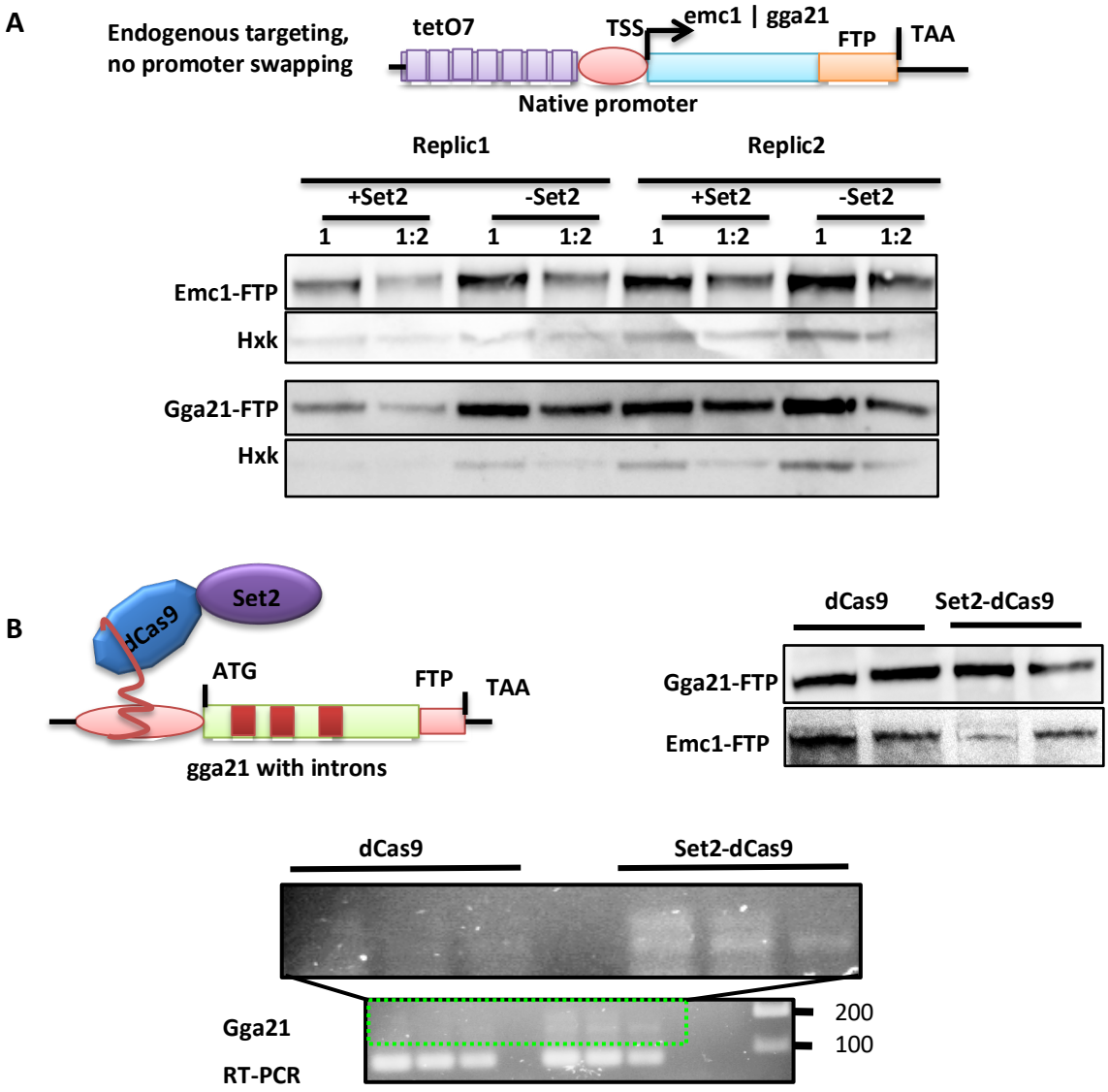



**Figure A.5. Tethering TetR-Set2 to native homologous TATA-less promoters doesn't affect protein signal of targeted genes. Targeting a single copy of Set2-dCas9 shows a possible impact on mRNA splicing. (A)** schematic representation for system of targeting homologous promoters (top) and wet Western blot analysis using PAP antibody (bottom) showing protein signal not affected by tethered TetR-Set2. **(B)** schematic representation for Set2-dCas9 targeting system (top left), wet Western blot analysis with anti-FLAG antibody (top right) showing protein signal not affected by tethered dCas9-Set2 and agarose gel analysis (bottom) for RT followed by PCR using RNA gga21 primers showing multiple PCR products in Set2-dCas9 but not dCas9 strain. PCR was done for 30 cycles, 20 µl reaction volume was loaded on 1.5% prestained agarose gel. Different dilutions were used for the protein extract to differentiate Western blot signals.

### A.3.5 TetR-Set2-targeted TATA*cyc1p*-driven genes have poly-adenylated transcripts and multiple transcription start sites

For further investigation, I decided to depend on the TetR-Set2 system with ectopic TRE and heterologous TATA*cyc1p* as results obtained by this system were consistent and reproducible. At this point, it was very important to make sure that the phenotype induced by TetR-Set2 isn't a manifestation of improper transcription initiation or defective RNA processing that may prevent efficient RNA translation. To exclude these possibilities, I investigated the features and characteristics of the transcripts of controlled genes. As a beginning, I wanted to examine the poly-A tail of these transcripts and determine the transcription start site or the most upstream 5' end of the transcripts. To examine the A-tail, I conducted reverse transcription using either oligo-dT or random primer mix followed by multiple standard PCRs using different primer pairs to examine the length of transcripts. Importantly, the reverse primer was a common primer specific for the FLAG sequence while different forward primers were designed to bind different sequences upstream the earliest known transcription start site (TSS) of TATA*cyc1p* (Figure A.6A). One of the primers binds exactly 15 bp upstream the earlies known TSS of TATA*cyc1p*. Successful PCR amplification using this primer would confirm the formation of full-length mRNA.

Surprisingly, using either oligo-dT or random primer mix for cDNA synthesis showed no difference in Ct values after RT-qPCR experiments for either H3 or lacZ genes (data not shown). This excludes that the abrogation of protein signal is a result for the absence of A-tail in mRNA. Moreover, PCR amplification for cDNA using multiple primer pairs to check the length of transcripts for TetR-Set2-targeted H3, produced amplicons of similar sizes for both tethered and non-tethered TetR-Set2 (Figure A.6A). As a control, I did the same experiment using the TetR-tup11 system to show that no PCR signal was detected for TetR-tup11-targeted H3 cDNA (Figure A.6A). Although this confirmed that the transcripts of TetR-Set2-targeted genes have full lengths and are not truncated, this wasn't enough to confirm identical TSS.

After confirming that the transcripts aren't truncated, I sought to determine the exact TSS of the targeted genes. This is required to exclude that the TSS doesn't exist far upstream or downstream the start codon, which would either form a long 5' UTR and decrease the competency of translation, or lose the start codon and disrupt the ORF, respectively. To address this, I applied the 5' RACE technique to the TetR-Set2 & lacZ system (which shows a striking

similarity in the RNA level). After RNA extraction and cDNA synthesis using lacZ specific primers, the cDNA was dC-tailed then extended using GI primer to generate second strand DNA (Figure A.6B). Remarkably, PCR amplification for the generated dsDNA produced a smeary product instead of a single band, indicating transcription initiation at different TSS and consistent with the scattered nature of transcription initiation at the TATA*cyc1p* **(Hahn *et al.*, 1985)**. This was shown for TetR-Set2-targeted and non-targeted lacZ (Figure A.6C). Next, I decided to clone 5' RACE products in *E. coli* for sanger sequencing, which confirmed the presence of multiple transcription start sites for the targeted and non-targeted lacZ. Importantly, the captured TSSs didn't show the formation of neither long UTRs or missing start codons. This eliminated the doubt that the abrogation of protein signal by TetR-Set2 is a consequence of defective transcription initiation or incompetent translation.

Recently, a mechanism for regulation of protein expression has been shown to occur through switching mRNA isoforms. Formation of 5' extended isoforms, also known as Long Undecoded Transcript Isoform (LUTI), can decrease the translation efficiency and the protein level compared to the canonical RNA **(Cheng *et al.*, 2018)**. To test this possibility, I checked the RNA level at the TATA*cyc1p* for the TetR-Set2-targeted and non-targeted TATA*cyc1p*-lacZ. Explicitly, the RNA level at the 5' end of TATA*cyc1p* was almost the same which excludes this possibility (Figure A.6D).

### A.3.6 TetR-Set2-induced abrogation of protein signal wasn't affected by deletion of ime4 m6A-methyl-transferase

The above observations suggest that a change in the RNA features rather than its level may be the reason for the absence of protein expression. TetR-Set2 may induce changes in RNA structure and modify it to either export- or translation-incompetent form. I speculated that Set2 may recruit an RNA methyltransferase that may edit the RNA and inhibit the translation. To check this hypothesis, I deleted ime4, the only m6A methyltransferase in *S. pombe,* to test whether this will eliminate the effect of TetR-Set2. Apparently, this didn't make a difference, as tethering TetR-Set2 to the TATA*cyc1p* was still able to abolish the protein signal of H3 (Figure A.7A). It's unlikely that the repression phenotype is a result of m6A RNA modification.

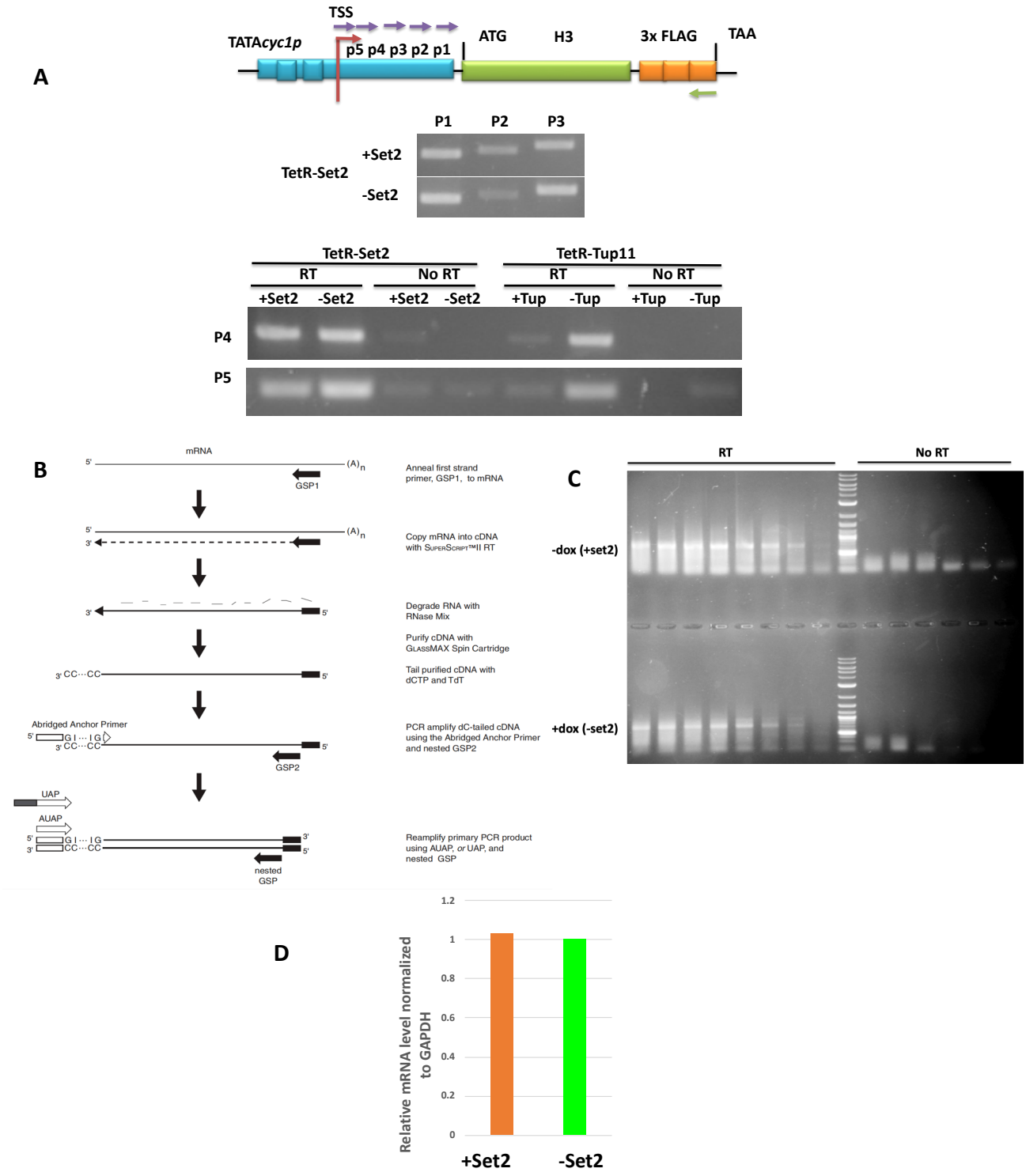



**Figure A.6. Transcripts of TetR-Set2-targeted TATA*cyc1p*-controlled genes are poly-adenylated and have multiple TSSs**. **(A)** schematic representation showing place of primers used for RT-PCR to detect length of H3-FLAG RNA (top) and agarose gel analysis showing successful amplification with all primers for TetR-Set2 targeted, but not TetR-Tup11-targetd H3. **(B)** outline of 5' RACE workflow adapted from ThermoFischer Scientific kit manual which I followed its principle. **(C)** agarose gel analysis for 5' RACE-PCR products, showing smeary product for LacZ. **(D)** RT-qPCR analysis showing similar promoter RNA level for TATA*cyc1p*-LacZ when targeted or un-targeted by TetR-Set2.

### A.3.7 TetR-Set2-induced effect can't be explained by either defective cap structure or action of the proteasome

Next. I guessed that H3K36me might recruit a pol II CTD kinase or phosphatase to phosphorylate or dephosphorylate, respectively, a specific pol II CTD residue that might prevent proper RNA capping and prevent binding of cap binding protein. To test this possibility, I investigated the effect of set2 deletion on phosphorylation levels of different Pol II CTD residues. For this, I used separate WT and *set2Δ* strains and checked the levels of Pol II phosphorylated Ser5, Ser2, Ser7, Tyr1 and Thr4 of the C-terminal domain of Rpb1 subunit using Western blotting. Again, I couldn't observe a remarkable difference in phosphorylation levels (A.7B). It's important to note that such expected changes may be very hard to detect with Western blotting, and a more dedicated method like ChIP is required to check the difference in the genomic enrichment of Pol II with different CTD phosphorylated residues, between the WT and the *set2Δ*.

To indirectly test the possibility that the cap structure may be the reason, I examined the levels of capped transcripts of targeted compared to non-targeted lacZ. For this, I carried out RNA immunoprecipitation (IP) using anti-m7G cap antibody, followed by RT-qPCR to check the difference in the levels of capped RNA. RT-qPCR analysis for IPed RNA, revealed that lacZ have similar capped-RNA levels in both conditions (Figure A.7B). This suggests that the cap structure isn't the reason behind the absence of protein signal.

One of the plausible explanations was that TetR-Set2 may induce a downstream pathway involving protein degradation by the proteasome. To test this possibility, I treated the culture overnight with MG132 to inhibit the endogenous proteasome machinery. Unexpectedly, this didn't restore the protein signal abolished by TetR-Set2 (Figure A.7D).

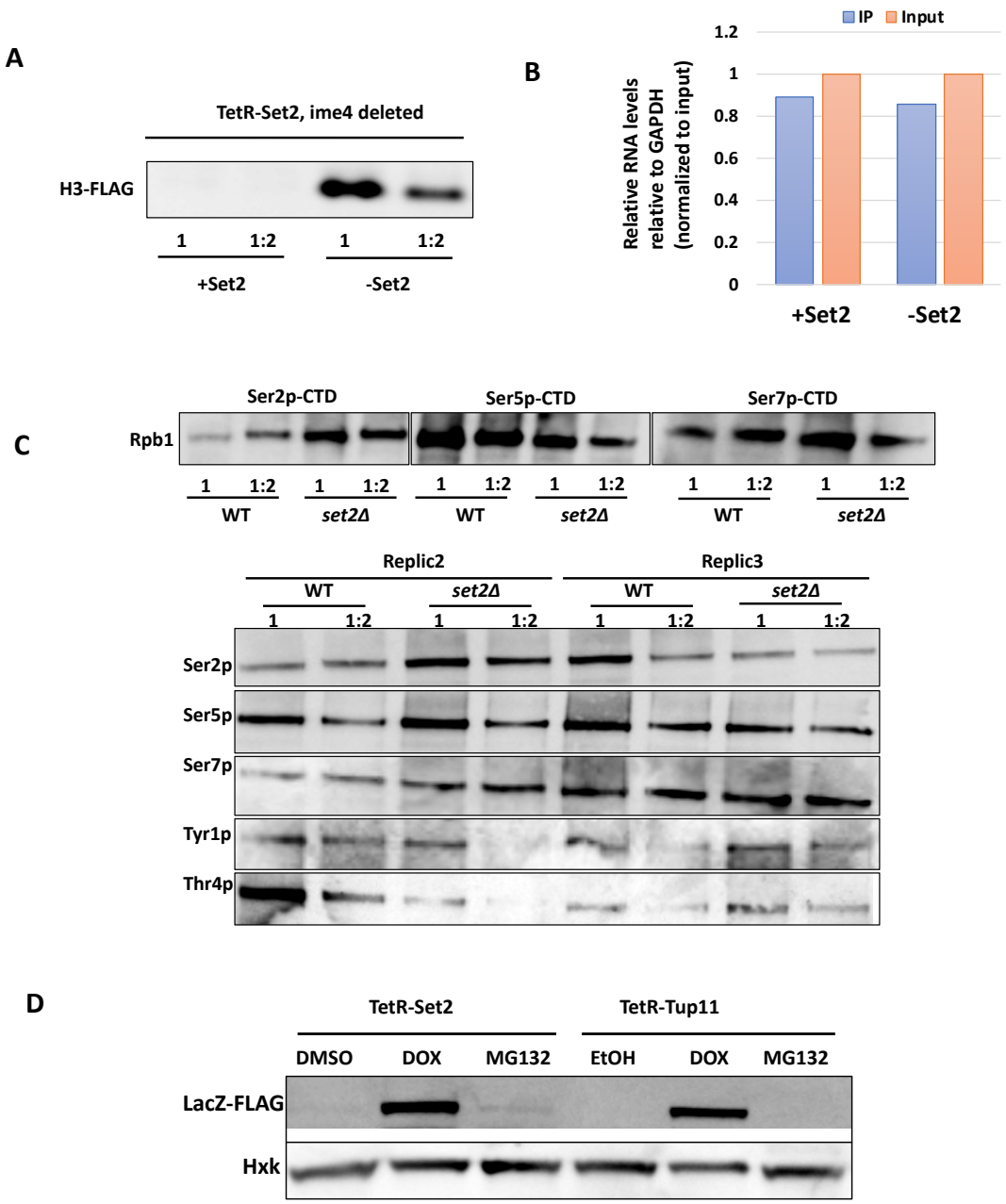



**Figure A.7. Absence of protein signal couldn't be explained by neither m6A modification or disrupted capping.** **(A)** RT-qPCR analysis for capped RNA IP using m7G antibody, showing comparable signals for capped RNA of lacZ. Ct values for lacZ were normalized to those of gpd3 then to the input. **(B)** semi-dry Western blot analysis showing that lacZ protein signal isn't detected after deleting ime4. Membrane was probed with HRP-conjugated anti-FLAG antibody and detected directly. Different dilutions were used for the total protein extract. **(C)** wet Western blot analysis showing no difference in levels of CTD phosphorylated Pol II Rbp1 subunit between WT and *set2Δ* yeast strains. Different dilutions were used for the protein extract to differentiate the signals. Samples used for the first row represent different biological replicates for the rest of the samples. Replic 1 and replic 2 represent different biological replicates. **(D)** Western blot analysis showing that protein signal wasn't restored after incubating culture overnight with 20 μM MG132 protease inhibitor.

### A.3.8 Abrogation of protein signal isn't induced by Set2, but by TetR alone

After extensive effort to understand this phenotype, I got doubtful that this phenotype is induced by Set2 itself. To clear these doubts, I planned to tether a mutant Set2 version, instead of the WT. For this, I mutated set2 at C261, a highly conserved residue in the SET domain, the catalytic methyltransferase domain of Set2 (Figure A.8), and used it for the same approaches above. For consistency reasons, I generated new strains with tetR-tup11, tetR-set2(WT) in parallel with the tetR-*set2*$^{C261A}$ strain. All the strains were generated to carry the same TRE with lacZ as a reporter gene. Tethering TetR-*Set2*$^{C261A}$ to TATA*cyc1p* still abrogated the protein signal of LacZ. In fact, abrogation of lacZ protein signal, that was restored by adding Dox, was shown for all newly generated strains (Figure A.8). Unfortunately, this means the abrogation of protein signal is induced by the TetR, but not Set2. Importantly, this Set2 mutation was shown to abolish H3K36me3 **(Landry *et al.*, 2003; Strahl *et al.*, 2002)**. However, recently, some mutations have been found to only affect the trimethylation but not the di- or the mono-methylation. Especially H3K36 di-methylation has been found to have a repressive affect similar to the trimethylation **(Hacker *et al.*, 2016)**.

To have a clear-cut evidence, I decided to use another negative control. For this, I used the DNA binding domain of TetR protein alone without fusing to neither Set2 nor Tup11. Unfortunately, I found that the TetR protein alone can also abolish the protein signal of lacZ (Figure A.8). Next, I studied the effect of tethering TetR alone on the RNA level of lacZ. I found that TetR doesn't decrease the RNA level of lacZ (Figure A.9). Unexpectedly, while using the new strains with tetR-tup11 as a control for no lacZ RNA transcription, I found that in the new strain, even the TetR-Tup doesn't affect the RNA level of lacZ, which was so confusing. Replacing lacZ with H3 in the same strain showed the same effect (Figure A.9). Obviously, none of the different protein-fusion systems seem to have an effect on the target RNA level (Figure A.9).

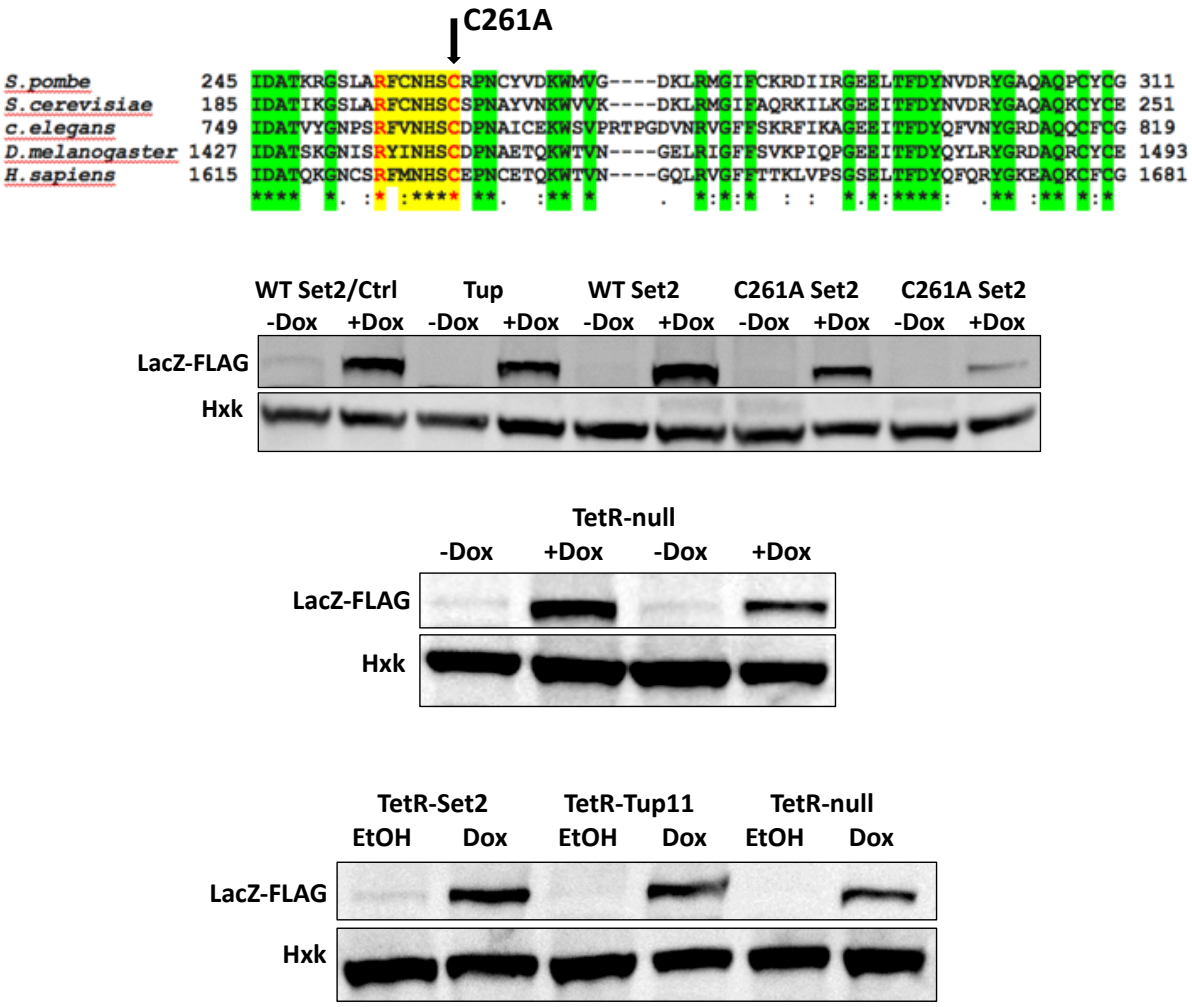



**Figure A.8. Abrogation of protein signal isn't induced by Set2, but TetR. (A)** ClustalW alignment for Set2 proteins from different organisms, showing highly conserved SET domain (yellow) and C261 residue that I mutated. **(B)** Western blot analysis showing protein signal of LacZ is abrogated with all tethered complexes and even with TetR alone (TetR-null).

These observations mean that in all cases, the TetR protein itself is the main responsible factor for repressing the protein signal of the reporter genes, and neither Set2 nor Tup has any role in this process. Surprisingly, only one specific strain that has the TetR-Tup and H3-FLAG and that previously showed that Tup11 can inhibit transcription and completely deplete H3 RNA (A.2.3) still shows the original results (Figure A.9). Clearly, in this strain, H3-FLAG construct was inserted in a locus that is completely different from the locus where both H3 and lacZ were inserted in the new strains. This means even the transcriptional repression effect of Tup11 is dependent on the locus (chromatin environment) of the targeted promoter.

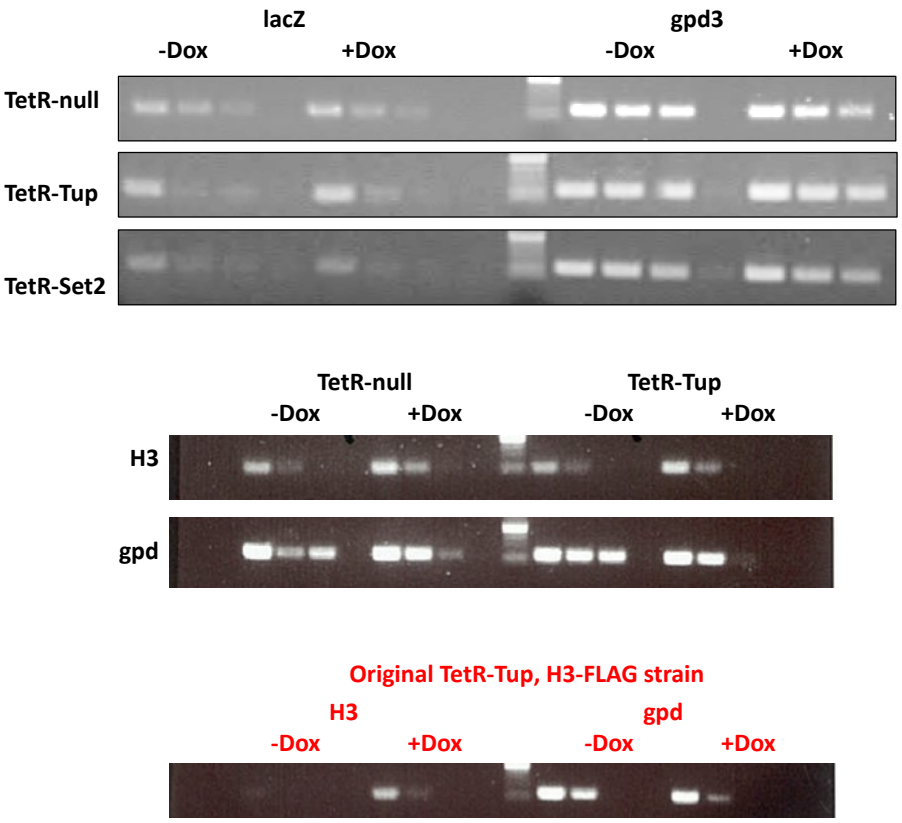



**Figure A.9. Transcription repression by Tup11 is dependent on the location of TRE response elements.** Agarose gel analysis after RT-PCR showing that none of the fusion proteins, including TetR-Tup11, affects RNA level of reporter gene. Only one strain with TetR-Tup and H3-FLAG gene inserted at a specific genomic locus shows depletion of H3 RNA by TetR-Tup.

## A.4 Discussion

Here I identified a novel artificial phenotype where mRNA is uncoupled to its protein expression, which contradicts with the central dogma of molecular biology. This phenotype is distinguished by abrogation of protein signal without a decrease in the RNA level of the target gene (Figure A.4). I identified this phenotype through targeting Set2, using the tet system, to a heterologous promoter driving variable reporter genes (Figure A.1). This phenotype was distinguished from the familiar phenotype where the protein signal disappears as a reason of absent transcription (Figure A.3). Interestingly, this phenotype was reliant on promoter type, but not the target gene, as it was linked to TATA but not native TATA-less promoters. Despite being very intriguing, I found that this phenotype is induced by the TetR repressor used for the fusion and tethering, but not Set2. I tried to figure the mechanism behind this phenotype in different ways.

The initial aim of this work was to understand the mechanism by which Set2-mediated H3K36 methyl-transferase suppresses cryptic promoters and represses pervasive transcription. Since Set2 suppresses cryptic promoters within gene bodies, I expected that if Set2 is recruited to a canonical promoter, it would block transcription initiation from this promoter. To approach this, I used different targeting systems to direct Set2 to different promoter systems and to study the response of different promoter types. First, I used the tet system to tether Set2 in a fusion protein complex (TetR-Set2) to a heterologous promoter-reporter gene system (TATA*cyc1p*-reporter gene). Tethering TetR-Set2 to the TATA*cyc1p* abrogated the protein signal of the controlled genes. Incredibly, the RNA levels of targeted genes were similar to those of the untargeted ones. While the RNA level didn't change at all for the targeted lacZ gene, it was dropped by only 25% for H3 gene, which still can't explain the absence of their corresponding proteins. I confirmed this phenotype by comparing against the regular phenotype produced by TetR-tup11 which shuts down transcription when tethered to the same promoters (Figure A.4).

At the beginning, I thought that this phenotype is imparted by Set2, so I wanted to understand the molecular requirements and factors determining this phenotype. First, I wanted to see whether this phenotype can be reproduced with other promoters of different types, so I inserted the tet operon repeats (tetO7) upstream the promoter region of some native genes to study their RNA and protein expression. Apparently, this new phenotype wasn't produced for these native genes as no change was detected for their protein signal (Figure A.5). This suggested that the phenotype is specific to TATA promoters only. Here it's important to stress that the native promoters are present at a genomic location different from that of the heterologous TATA*cyc1p* system. For this, it should be kept in mind that this phenotype might be imparted by specific chromatin features (the chromatin environment of the locus).

In parallel with the tet system which enabled tethering multiple copies of Set2, I used dCas9 to tether a single Set2 molecule to the native promoters of the same non-essential genes, to study a foreseeable differential effect for one Set2 molecule. Although this didn't show an impact on the protein signal, an impact on splicing was revealed by the production of different amplicons after RT-PCR (Figure A.5). These observations are in line with a known role for Set2 in pre-mRNA splicing **(McDaniel & Strahl, 2017; Meers *et al.*, 2017; Sorenson *et al.*, 2016; Wagner & Carpenter, 2012; K. Zhu *et al.*, 2017)**, but still need to be confirmed before further investigation.

I carried out extensive trials to understand this phenotype. I examined the A-tail and the TSS of the RNA using 5' RACE followed by cloning and sequencing. Clearly, I found that the targeted-genes transcripts have a poly-A tail and multiple transcript start sites similar to the control (un-targeted genes). I tested the RNA level at the promoter using RT-qPCR and the levels of m7G-capped RNA using IP followed by RT-qPCR which didn't reveal a difference (Figures A.6 & A.7). These findings exclude that this phenotype is a consequence of aberrant transcription initiation, defective RNA processing or improper capping. To test whether this phenotype is induced by m6A RNA modification, I deleted the m6A methyltransferase ime4, which didn't affect the phenotype (Figure A.7). Importantly, inhibiting the endogenous proteasome using MG132 didn't restore the protein signal. This suggests that the RNA may not be translated at all (Figure A.7).

After extensive trials, I realized that Set2 isn't responsible for the absence of the protein signal, but definitely the TetR alone which was able to show the same phenotype (Figure A.8). In retrospect, two early studies have shown that tethering Set2 to either CYC1 **(Strahl *et al.*, 2002)** or GAL4 promoter **(Landry *et al.*, 2003)** in *S. Cerevisiae* represses transcription and suppresses protein expression of controlled genes. Inconsistently, however, the transcription repression effect by Set2 was inferred from the enzymatic activity of the controlled genes proteins (GAL 4 and LacZ) not from the RNA level. For this specific reason, I thought that the novel phenotype which I identified here may be really induced by Set2. In addition, the same tet system has been frequently used for targeting different epigenetic modifiers to specific response systems **(Audergon *et al.*, 2015; Jih *et al.*, 2017; Ragunathan *et al.*, 2015; X. Wang & Moazed, 2017; R. Yu *et al.*, 2018)**. For these reasons, at the beginning, I didn't have any doubt that this effect is independent of Set2. Regardless, I noticed that tethering TetR-Set2 slightly increases the RNA levels of controlled genes (Figures A.4, A.6 & A.7).

However, it was confusing to see that the TetR-Tup also exhibited the same new phenotype when the response elements (reporter gene) were inserted at the same locus used for tetR-set2 strains response elements (Figure A.9). Surprisingly, this suggests that promoter regulation of gene expression is dependent on the genomic location within chromatin. These observations show that the chromatin environment may induce epigenetic control that is superior to the promoter type and DNA sequence. These results are consistent with the "domain hypothesis of the eukaryotic genome organization.", according to which, different mechanisms including chromatin domains coordinate gene regulatability **(Recillas-Targa & Razin, 2001)**.

For instance, chromatin domain organization separates constitutively expressed form developmentally regulated genes in *C. elegans* **(Evans *et al.*, 2016)**. It's important to note that the response elements containing the heterologous TATA*cyc1p* have the same structure, the same marker gene and the same sequences, but, only the location of genomic integration changes.

Considering the bright side of this finding, it's possible that the effect of Set2 and its mechanism of action involving H3K36 methylation have been masked, since the target promoters are inserted in a non-regulatable genomic domain. In order to understand Set2 function, the response elements have to be inserted in a locus that allows epigenetic regulation by histone modifications or other chromatin factors.

It's not really clear how TetR can induce this phenotype knowing that only the DNA-binding domain of this protein is used. This system is widely used and has never been reported to show any interference with gene expression. Probably, it can promiscuously bind to the RNA and block its transport or export which needs to be investigated. It may nonspecifically bring other factors that may induce this phenotype which doesn't seem to be co-transcriptional as the RNA level isn't affected. It should be kept in mind that this phenotype doesn't seem to be universal but promoter/chromatin-domain dependent. The reason is that it was produced only with the TATA*cyc1p* but not the TATA-less native promoters.

## A.5 Conclusion

I identified a novel *in vivo* phenotype induced by tethering bacterial TetR to a heterologous promoter system in fission yeast. Tethering TetR to this promoter system doesn't affect the RNA level, but abolishes the protein signal of the reporter genes. This phenotype is dependent on promoter type (TATA or TATA less) and probably genomic location of target system. These observations don't confirm or refute a role for Set2 that showed the same phenotype when fused with TetR (TetR-Set2) and targeted to the same promoter system. This means a neutral targeting system should be used instead of the tet system. I have no evidence to conclude whether this is a completely artificial, or whether it can be a natural unknown mode of gene regulation. Also, I'm not aware of any applications where the RNA is allowed to be transcribed but not translated. Can this be a defense system that targets exogenous transcription units?!

## A.6 Prospective

It is possible that the TetR can bind non-specifically to some factors associated with this TATA promoter, or associated with the genomic location (chromatin domain) of this promoter to induce this effect. A site-specific immunoprecipitation followed by mass spectrometry may be required to reveal factors possibly involved in this phenotype. I think a more precise method such as northern blotting, is required to check the full length of targeted genes transcripts compared to the untargeted ones. It may be insightful to study the export and the localization of the affected reporter transcripts and test their interaction with the ribosomes.

## A.7 Materials and methods

### A.7.1 Selection and maintenance of yeast and bacterial cells

For general purposes, *S. pombe* cells were grown at 30 °C on solid or in liquid YEA rich media. For auxotrophic selection, yeast cells carrying antibiotic marker genes were selected on solid YEA containing the corresponding antibiotic (200 µg/ml Nourseothricin, 200 µg/ml Geneticin (G418) or 400 µg/ml Hygromycin B). For prototrophic selection, yeast cells were selected on solid SDC media containing all supplements except the selective amino acid(s). After confirming the right genotype by PCR, yeast cells were maintained in non-selective conditions. DH5α *E. coli* cells (Bioline) were grown at 37 °C on solid or in liquid LB media containing 100 µg/ml Ampicillin for selection of cells containing the target plasmid. Media and yeast strains used in this study are listed in Table A.1 and Table A.2, respectively.

### A.7.2 Yeast and bacterial cell transformation

Freshly growing yeast cells were inoculated overnight in rich YEA medium to reach $OD_{600}$ of 0.4 -0.8. Yeast cells were harvested in 50 ml falcon by centrifugation at 2600 rpm at room temperature for 2 min. After removing supernatant, cells were washed with 20 ml autoclaved water and centrifuged at 2600 rpm at room temperature for 2 min. Supernatant was removed and cells were resuspended in 1 mL transformation solution 1 (0.1 M LiAc, 0.1 M TE) and transferred to an Eppendorf tube. Cells were pelleted by centrifugation at 2600 rpm for 2 min. After discarding supernatant, cells were resuspended in in 250 µl transformation solution 1 and kept on ice. To each 50 µl of conditioned cells was added 300 µl transformation solution 2 (0.1 M LiAc, 0.1 M TE, 40% PEG), 20 µl herring sperm (sigma), 5-20 µl DNA (1 to 2 µg). Contents were mixed by gently vortexing before incubating on a turning wheel for 30

to 45 min at 30°C. Cells were heat shocked at 42°C for 12-20 min before adding 1 ml YEA and centrifuging at 2600 rpm for 2 min at room temperature. Supernatant was removed, cells were resuspended in 1 mL YEA and incubated on a turning wheel for 2-4 hours. Transformed cells were pelleted by centrifugation at 2600 rpm, room temperature, 2 min before resuspending in 50-100 ml YEA and re-streaking on the corresponding selective plate. Transformed cells were left to grow at 30°C for 2 to 4 days, and single colonies were picked and streaked on selective plates.

DH5α *E. coli* cells were transformed following manufacturer manual; after thawing, cells were incubated with 2 µl of plasmid DNA for 30 min before heat shocking at 42°C for 30 sec. Cells were kept on ice for 2 min before adding 950 µl SOC medium and allowing to grow at 37 °C for 60 min with shaking at ~250 rpm. 100 µl transformation mix was plated on LB agar plates containing appropriate antibiotic and incubated overnight at 37°C.

### A.7.3 PCR colony check and high-fidelity PCRs

For standard PCRs, Promega *Taq* polymerase was used for amplification. Before colony check for selected or newly generated yeast strains, yeast cells were collected from a freshly growing plate using white loop and mixed in 5 µl of Zymolyase mix (1.2 M Sorbitol, 0.5 mg/ml Zymolyase 100 T [Nacalai Tesque], 100 mM $NaH_2PO_4$) for cell wall digestion (spheroblasting) by incubating in a PCR machine with program; 37 °C/15 min, 98 °C/2 min, 4 °C/∞. After diluting the spheroblast in 100 µl H2O, 2 µl were used as a template for 20 µl *Taq* PCR reaction volume. For colony check primers, one primer was designed to bind to a unique sequence in the inserted cassette and the second was designed to bind to a genomic sequence outside the homology sequence used for integration. *E. coli* single-cell colonies were scrapped using white loop and directly mixed in PCR reaction mix.

For high fidelity PCR amplifications purposes including cloning, sequencing, and genome editing, a home-made high fidelity Phusion Polymerase was used for amplification following ThermoFischer Scientific specifications for PCR program and conditions. GC buffer and DMSO were added for amplification of high GC DNA amplicons.

### A.7.4 Genome editing: gene deletion, epitope tagging and sequence insertion

Unless otherwise stated Gibson assembly method was adapted for all plasmid cloning and cassette generation. Genome editing of fission yeast was performed using PCR amplified DNA fragments or linearized vectors and adopting the long homology sequence principle. All cassettes were designed to have a resistance marker gene plus 5' and 3' flanking sequences, each of around 500 bp-length, homologous to those surrounding the target integration site. Insert sequence and/or the marker gene were amplified from a plasmid while the 500 bp homology sequence was amplified from genomic DNA. The generated fragments were assembled together using home-made Gibson assembly kit following NEB protocol recommendations. The final assembled construct was amplified then purified prior to transfecting in yeast. Deletion cassettes were designed to replace the coding sequence by including two flanking regions amplified upstream the ATG and downstream the TAA sequence. C-terminal FTP tagging was done following a similar principle; a DNA fragment containing the selection marker gene and the tagging sequence were amplified from the corresponding plasmids then assembled with two 500 bp fragments homologous to the integration site sequence. After Gibson assembly, the whole fragment was amplified and transfected in yeast to be inserted directly before the stop codon of the target gene. For inserting the tet response elements for *in situ* control of gene expression, a DNA fragment of tetO7 repeats (without the TATA*cyc1p*) and the selection marker gene were amplified from plasmids and ligated to two ~500 bp homology sequences using Gibson assembly, to be finally inserted upstream the promoter of the target genes. Plasmids and oligos used for genome editing are listed in Table A.3 and Table A.4, respectively.

### A.7.5 Cloning and design for the tet Trans-Suppressor system (tTS)

Strains containing the tTS system were generated on two steps to integrate either the trans-suppressor (tetR-set2, tetR-tup or tetR-null) or the tet response elements (tetO7-TATA*cyc1p*-reporter gene) constructs. The suppressor construct sequence of tetR-set2, tetR-tup or tetR-null was first cloned in a plasmid that was linearized and transfected in *ura4-D18* yeast strains for integration at the ura4 locus. pDM291-tetR-tup11Δ70 plasmid was digested to excise the tup and either clone set2 to generate tetR-set2 or completely get rid of suppressor sequence to generate TetR-null. The sequence of tet response elements containing the tagged reporter gene was also cloned in a plasmid that was digested to produce a long fragment with

flanking homology sequences for integration at leu1 locus. For *in situ* suppression of endogenous promoters, the suppressor construct was integrated the same way above, while the tet response elements (without TATA*cyc1p*) were amplified to be inserted upstream the endogenous promoter of the target gene which was finally C-terminally FTP-tagged. Plasmids and oligos used for this system are listed in Table A.3 and Table A.4, respectively.

For final experiments, yeast cells were inoculated and allowed to grow without dox then split into two halves before overnight cultivation. 10 µM doxycycline was added for one of the cultures for overnight induction (tetR-release from promoter). Finally, equal number of cells were harvested from each culture to be processed either for Western blotting or RNA extraction.

### A.7.6 Cloning and design for dCas9 and guide RNAs

dCas9 sequence was amplified from plasmid pTDH3-dCas9 (#242) to be cloned in pFA6a-5'Leu1-kanMX6-TATA*cyc1p*-3'Leu1 (#340) downstream the marker gene after excising the TATA*cyc1p* sequence. For generation of N-terminal fusion of set2 to dCas9, set2 was amplified from plasmid pDM291-tetR-set2 using primers that add a triglycine linker (L3) required for cloning set2-L3 upstream dCas9 to have one open reading frame. For the final step of cloning, gRNA fragment with a separate promoter, scaffold RNA and terminator, was cloned in both of the dCas9 and set2-dCas9 plasmids in a separate ORF. This was done on two steps; first, gRNA scaffold plasmid was used as a template for generation of two short fragments. Each fragment has the specific gRNA sequence (added using amplification primers) to be in the middle as homology sequence for Gibson assembly. After GA, the final product is amplified using nested PCR to generate one DNA fragment. The final plasmid was digested to integrate the whole construct at the leu 1 locus. Plasmids and oligos used for cloning are listed in Table A.3 and Table A.4, respectively.

**Table A.1. List of media used and their components.**

| Medium | components |
|---|---|
| YEA | 0.5% (w/v) yeast extract (MP Biomedicals), 3% (w/v) glucose (Sigma), 0.8 mM adenine (Sigma), 2% (w/v) agar (Sigma) for solid medium, pH 5.5. |
| SDC-X | 2% (w/v) glucose (Sigma), 0.67% (w/v) yeast nitrogen base without amino acids (Formedium), 0.162% (w/v) CSM-X (Formedium), 2% (w/v) agar (Sigma) for solid medium, pH 5.5 |

| | |
|---|---|
| LB | 0.5% (w/v) yeast extract (MP Biomedicals), 1% (w/v) tryptone (MP Biomedicals), 0.5% (w/v) sodium chloride (Sigma), 1.65% (w/v) agar (Sigma) for solid medium, pH 7.2 |

**Table A.2. List of *S. pombe* strains used in this study**

| Strain no. | Strain title | Genotype | Reference |
|---|---|---|---|
| P20 | WT | *h+, leu1-32, ura4 DS/E, ade6-216, HIS+, Otr1R::Ura4 (SphI)* | Lab strain |
| P70 | set2Δ::KanMx | *h+, leu1-32, ura4 DS/E, ade6-216, HIS+, Otr1R::Ura4 (SphI), set2::KanMx,* | Lab strain |
| P312 | set2Δ:hphNT1/1 | *h+, leu1-32, ura4 DS/E, ade6-216, HIS+, Otr1R::Ura4 (SphI), set2Δ::hphNT1,* | Lab strain |
| P344 | WT | *h+, HIS, leu1-32, ade6-210, ura4-D18* | Bioneer V2-33-H11 |
| P419 | WT | *MatMsmt0, leu1-32, ura4-D18, ade6- 216, his2* | Lab strain |
| P1075 | WT tup11Δ70-tet repressor /1 | *MatMsmt0, leu1-32, ura4-D18, ade6- 210, tetR-tup11Δ70::ura4, his2* | Lab strain |
| P1242 | set2Δ::hphNT1 | *Mat1Msmt0, his2, leu1-32, ade6-216, ura4 DS/E or ura4-D18?* | Lab strain |
| P1558 | WT histone turnover strain - alternative | *h+, HIS+, leu1-32, ade6- 210, Otr1R::Ura4, ura4-D18, tup11Δ::Ura4, hphNT1::TetO7::hht1-6xFlag::Nat NT2, cdc25-22 (ts)* | Lab strain |
| F2931 | TetR-set2 | *CMVp-tetR-set2::Ura4, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F2942 | TetR-set2, tetO7-cyc1p-hht1-3xFLAG | *CMVp-tetR-set2::Ura4; kanMX:: tetO7-CYC1p-hht1-3xflag, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F2943 | TetR-set2, tetO7-cyc1p-hht1-3xFLAG | *CMVp-tetR-set2::Ura4; kanMX:: tetO7-CYC1p-hht1-3xflag, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F2979 | dCas9-emc1gRNA | *dCas9-emc1gRNA::KanMX, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F2983 | dCas9-gga21gRNA | *dCas9-gga21gRNA::KanMX, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F2987 | Set2-dCas9-emc1gRNA | *Set2-L3-dCas9-emc1gRNA::KanMX, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F2991 | Set2-dCas9-gga21gRNA | *Set2-L3-dCas9-gga21gRNA::KanMX, h+, HIS ,leu1-32, ade6-210, ura4-D18* | This study |
| F3011 | ime4Δ::KanMX4 #1 | *h+, HIS, leu1-32, ade6-210/216, ura4-D18* | Bioneer V2-22-A07 |
| F3017 | dCas9-emc1gRNA, emc1-CtFTP | *emc::CtFTP-NatNT2, 5'leu1::KanMx-dCas9-emc1gRNA::3'leu1, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3018 | dCas9-gga21gRNA, gga21-CtFTP | *gga21::CtFTP-NatNT2, 5'leu1::KanMx-dCas9-gga21gRNA::3'leu1, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |

| | | | |
|---|---|---|---|
| F3019 | set2-dCas9-emc1gRNA, emc1-CtFTP | *emc::CtFTP-NatNT2, 5'leu1::KanMx-set2-L3-dCas9-emc1gRNA::3'leu1, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3020 | set2-dCas9-gga21gRNA, gga21-CtFTP | *gga21::CtFTP-NatNT2, 5'leu1::KanMx-L3-dCas9-gga21gRNA::3'leu1, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3021 | TetR-set2, tetO7-cyc1p-lacZ-3xFLAG | *5'leu1::kanMX-tetO7-CYC1p-lacZ-3xflag::3'leu1, CMVp-tetR-set2::Ura4, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3023 | TetR-set2, tetO7-cyc1p-gga21-3xFLAG | *5'leu1::kanMX-tetO7-CYC1p-genomic gga21-3xflag::3'leu1, CMVp-tetR-set2::Ura4, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3025 | TetR-set2, emc1-CtFTP | *emc1::CtFTP-NatNT2, CMVp-tetR-set2::Ura4, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3027 | TetR-set2, gga21-CtFTP | *gga21::CtFTP-NatNT2, CMVp-tetR-set2::Ura4, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3032 | TetR-set2, tetO7-cyc1p-hht1-3xFLAG, ime4Δ | *ime4Δ::hphNT1, CMVp-tetR-set2::Ura4; 5'leu1::kanMX-tetO7-CYC1p-hht1-3xflag::3'leu1, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3033 | TetR-set2, tetO7-cyc1p-spliced gga21-3xFLAG | *5'leu1::kanMX-tetO7-CYC1p-RT gga21-3xflag::3'leu1, CMVp-tetR-set2::Ura4, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3035 | TetR-set2, emc1-CtFTP, tetO7-emc1p | *5'emc1::kanMX-tetO7::emc1p, emc1::CtFTP-NatNT2, CMVp-tetR-set2::Ura4, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3037 | TetR-set2, gga21-CtFTP, tetO7-gga21p | *5'gga21::kanMX-tetO7::gga21p, gga21::CtFTP-NatNT2, CMVp-tetR-set2::Ura4, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3039 | tetO7-cyc1p-lacZ-3xFLAG | *5'leu1::kanMX-tetO7-CYC1p-lacZ-3xflag::3'leu1, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3060 | tetO7-cyc1p-lacZ-3xFLAG, TetR-Tup11Δ70 | *h+, HIS, leu1-32, ade6-210, ura4-D18, 5'leu1::kanMX-tetO7-CYC1p-lacZ-3xflag::3'leu1, CMVp-tetR-Tup11Δ70::Ura4* | This study |
| F3062 | tetO7-cyc1p-lacZ-3xFLAG, TetR-Set2 | *h+, HIS, leu1-32, ade6-210, ura4-D18, 5'leu1::kanMX-tetO7-CYC1p-lacZ-3xflag::3'leu1, CMVp-tetR-Set2::Ura4* | This study |
| F3064 | tetO7-cyc1p-lacZ-3xFLAG, TetR-C261A Set2 | *h+, HIS, leu1-32, ade6-210, ura4-D18, 5'leu1::kanMX-tetO7-CYC1p-lacZ-3xflag::3'leu1, CMVp-tetR- C261A Set2::Ura4* | This study |
| F3066 | tetO7-cyc1p-lacZ-3xFLAG, TetR-null | *h+, HIS, leu1-32, ade6-210, ura4-D18, 5'leu1::kanMX-tetO7-CYC1p-lacZ-3xflag::3'leu1, CMVp-tetR only::Ura4* | This study |
| F3072 | tetO7-cyc1p-hht1-3xFLAG | *kanMX:: tetO7-CYC1p-hht1-3xflag, h+, HIS, leu1-32, ade6-210, ura4-D18* | This study |
| F3073 | tetO7-cyc1p-hht1-3xFLAG, TetR-null | *h+, HIS, leu1-32, ade6-210, ura4-D18, kanMX:: tetO7-CYC1p-hht1-3xflag, CMVp-tetR only::Ura4* | This study |

| | | | |
|---|---|---|---|
| F3075 | tetO7-cyc1p-hht1-3xFLAG, TetR-Tup11Δ70 | *h+, HIS, leu1-32, ade6-210, ura4-D18, kanMX:: tetO7-CYC1p-hht1-3xflag, CMVp-tetR-Tup11Δ70::Ura4* | This study |
| F3077 | TetR-Tup11Δ70, tetO7-cyc1p-LacZ-3xFLAG | *h+, HIS, leu1-32, ade6-210, ura4-D18, CMVp-tetR-Tup11Δ70::Ura4, kanMX:: tetO7-CYC1p-LacZ-3xflag* | This study |

**Table A.3. List of plasmids used in this study.**

| Plasmid no. | Plasmid name and structure layout | Application | Reference |
|---|---|---|---|
| 1 | pFA6a-natNT2 | Deletion cassettes with nourseothricin marker gene | Hurt lab #3191 |
| 2 | pFA6a-hphNT1 | Deletion cassettes with hygromycin B marker gene | Hurt lab #3190 |
| 3 | pFA6a-kanMX | Deletion cassettes with geneticin marker gene | Hurt lab #4512 |
| 12 | pFA6a-FTpA-TCYC1-natNT2 | C-terminal tagging of proteins with Flag-TEV-Prot. A | Hurt lab #3486 |
| 91 | pDM291-tetR-tup11Δ70. | Integrating tetR-tup11Δ70 & Base for tetR-set2 plasmids | Euroscarf. #30677 |
| 122 | pFA6a-5'Leu1-kanMX6-tetO7-TATA*cyc1p*-3'Leu1 | Base for tet response elements plasmids. | Lab plasmid |
| 241 | pU6-sgGAL4-1 | Cloning and sequencing 5' RACE products | Addgene plasmid #46915 |
| 242 | pTDH3-dCas9 | PCR amplification of dCas9 for subcloning. | Addgene #46920 |
| 243 | pSNR52-sgTEF1 | PCR Amplification of gRNA scaffold | addgene plasmid #46922 |
| 324 | pDM291-tetR-set2 (BamHI mutated). | Integrating tetR-set2 construct | This study |
| 326 | pFA6a-5'leu1–kanMX-tetO7-TATA*cyc1p*-H3.1-3xFLAG-3'leu1. | Integrating tet response elements with H3 gene. | This study |
| 327 | pDM291-tetR-set2(WT). | Integrating tetR-set2 construct. | This study |
| 328 | pMC1871 | PCR amplification of lacZ sequence | Ed Hurt |
| 340 | pFA6a-5'Leu1-kanMX6(mutNcoI)-tetO7-TATA*cyc1p*-3'Leu1. | Base for generating set2-dCas9 fusion plasmids. | This study |
| 341 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-dCas9-adhT-3'Leu1. | dCas9 cloned, base for fusing N-terminal set2 or only gRNA. | This study |
| 342 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-NTset2-L3-dCas9-adhT-3'Leu1. | set2-dCas9 cloned, base for cloning gRNA for different target genes. | This study |
| 343 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-dCas9-adhT-pSNR52-intron-emc1gRNA-Scaffold-SUP4t-3'Leu1. | Integration of control dCas9 with emc1 gRNA | This study |
| 344 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-dCas9-adhT-pSNR52-intron-gcn1gRNA-Scaffold-SUP4t-3'Leu1. | Integration of control dCas9 with gcn1 gRNA | This study |

| | | | |
|---|---|---|---|
| 345 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-dCas9-adhT-pSNR52-intron-gga21gRNA-Scaffold-SUP4t-3'Leu1. | Integration of control dCas9 with gga21 gRNA | This study |
| 346 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-dCas9-adhT-pSNR52-intron-ppk1gRNA-Scaffold-SUP4t-3'Leu1. | Integration of control dCas9 with ppk1 gRNA | This study |
| 347 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-NTset2-L3-dCas9-adhT-pSNR52-intron-emc1gRNA-Scaffold-SUP4t3'Leu1. | Integration of set2-dCas9 with emc1 gRNA | This study |
| 348 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-NTset2-L3-dCas9-adhT-pSNR52-intron-gcn1gRNA-Scaffold-SUP4t-3'Leu1. | Integration of set2-dCas9 with gcn1 gRNA | This study |
| 349 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-NTset2-L3-dCas9-adhT-pSNR52-intron-gga21gRNA-Scaffold-SUP4t-3'Leu1. | Integration of set2-dCas9 with gga21 gRNA | This study |
| 350 | pFA6a-5'Leu1-kanMX6(mutNcoI)-pTDH3-NTset2-L3-dCas9-adhT-pSNR52-intron-ppk1gRNA-Scaffold-SUP4t-3'Leu1. | Integration of set2-dCas9 with ppk1 gRNA | This study |
| 370 | pFA6a-5'leu1–kanMX6-tetO7-TATA*cyc1p*–lacZ-3xFLAG-3'leu1. | Integrating tet response elements with lacZ gene. | This study |
| 371 | pFA6a-5'leu1-kanMX-tetO7-TATA*cyc1p*-gga21(full)-3xFLAG-3'leu1. | Integrating tet response elements with full-length gga21 gene. | This study |
| 372 | pFA6a-5'leu1-kanMX-tetO7-TATA*cyc1p*-gga21(intronless)-3xFLAG-3'leu1. | Integrating tet response elements intron-less gga21 gene. | This study |
| 373 | pDM291-tetR-set2$^{C261A}$(mutBamHI). | Integrating tetR-set2$^{C261A}$ construct | This study |
| 374 | pDM291-TetR. | Integrating TetR alone | This study |

**Table A.4. List of oligos used for tagging, deletions and insertions, and dCas9 and set2 cloning.**

| No. | Name | Sequence | Application |
|---|---|---|---|
| A63 | GA.pTDH3.Fw | GACGAGGCAAGCTaaacacccgggCTAGTCAGTTCGAGTTTATCATTATC | dCas9 amplification into two fragments and Gibson assembly cloning in #122. |
| A64 | GA.dCas9.Rv | TCGTTCTCCTCATTGTCCAG | |
| A65 | GA.dCas9.Fw | GGACAATGAGGAGAACGAGG | |
| A66 | GA.adh1T.Rv | GCTAAgtacgctgcaggtcgacgaactGCTAGCtcttaACCGGTGAAATGGGGAGCGATTTGC | |
| | | | |
| A88 | GA.set2.pTDH3.Fw. | tGGAATTAGATCTCGCCACCATGCAGACGGCATCATCTCTTTCTG | Set2 amplification and Gibson assembly cloning in #341 |
| A89 | GA.set2.L3.dCas9.Rv. | TCCGATAGAATACTTCTTGTCtgatccacctgagcctccTGATCCGCCACCAGCAGCTTTTTTCGGGGATTC | |
| | | | |
| A96 | Amp.pSNR52-gRNA-scaf.Fw. | TCTTTGAAAAGATAATGTATGATTATGC | Used with # A100 or A104 |
| A97 | Amp.pSNR52-gRNA-scaf.Rv. | TATAGTCCTGTCGGGTTTCG | Used with A101 or A105 |

| | | | |
|---|---|---|---|
| A98 | GA.pSNR52.dCas9vec.Fw. | CAAATCGCTCCCCATTTCATCTTTGAAAAGATAATGTATGATTATGC | Final amplification of gRNA and scaffold for GA cloning. |
| A99 | GA.gRNAscaf.dCas9vec.Rv. | gacgaactGCTAGCtcttaACCGGTAGACATAAAAAACAAAAAAAGCACC | |
| A100 | GA.emc1gRNA.Rv. | ACACTGTAAACTTAGCAATCAAGATCATTTATCTTTCACTGCG | |
| A101 | GA.emc1gRNA.Fw | TTGATTGCTAAGTTTACAGTGTTTTAGAGCTAGAAATAGCAAGTTAAA | |
| A104 | GA.gga21gRNA.Rv. | ACCATGCTTTTTCAACGATATAGATCATTTATCTTTCACTGCG | |
| A105 | GA.gga21gRNA.Fw. | TATATCGTTGAAAAAGCATGGTTTTAGAGCTAGAAATAGCAAGTTAAA | |
| | | | |
| A153 | Amp.Intg.5'emc1.Fw | GCATTCAAAGGAGGAACATC | To assemble tetO7 repeats to two 500 bp homology sequences and the 5' end of emc1 and integrate in yeast |
| A154 | Intgr.tetO7.emc1p.Rv | gagctcgaattcatcgatgaCGTTTAGCAAACGTACTAAATGC | |
| A155 | Intgr.tetO7.emc1p.fw | GTACCCtatggCATGCATGGAATGTACAGGTGTTTCACAGC | |
| A156 | Amp.Intg.5'emc1.Rv | ATCACATCTCCGTTTGATGC | |
| | | | |
| A160 | Amp.intg.3'emc1.Fw | AATGATTACAGGCCCTATGC | For C-terminal FTP tagging of emc1 |
| A161 | Tag.FTP.Ct-emc1.Rv | CGTCATGGTCCTTGTAGTCATTGTACCATTTGGTATTAAGCTG | |
| A162 | Tag.FTP.Ct-emc1.Fw | gagctcgaattcatcgatgaTAAAAACACCATGGAGAATAAAAATATAC | |
| A163 | Amp.intg.3'emc1.Rv | CTAGTGTCGATGGTACTGTCG | |
| | | | |
| A165 | Amp.del.ime4.Fw | GACGTTAAGGCGAGATAAGG | Generation of ime4 deletion cassette and integration in yeast |
| A166 | Del.ime4.hphNT1.Rv | GCGTACGAAGCTTCAGCTTAAAAATTACACGAATACCGAAC | |
| A167 | Del.ime4.hphNT1.Fw | GAGCTCGAATTCATCGATGATGGAAAATGCTACATTGTTCG | |
| A168 | Amp.del.ime4.Rv | GTAATGAAGGGATAGCAGCC | |
| A172 | Amp.Intg.5'gga21.Fw | TAATCAATTGACCGTCCCTC | Integrate tetO7 repeats upstream gga21 promoter |
| A173 | Intgr.tetO7.gga21p.Rv | gagctcgaattcatcgatgaTTTATTGCTGTAGCTAGTGTCG | |
| A174 | Intgr.tetO7.gga21p.Fw | GTACCCtatggCATGCATGCTGCGGATCAAAAATATGGC | |
| A175 | Amp.Intg.5'gga21.Rv | CAGCAACAAAGCAATTGTAGG | |
| A176 | Amp.intg.3'gga21.Fw | TACTGAGGAACCTGCAGTCC | C-terminal FTP tagging for gga21 |
| A177 | Tag.FTP.Ct-gga21.Rv | CGTCATGGTCCTTGTAGTCCAGCGGAAGATGCGATTC | |
| A178 | Tag.FTP.Ct-gga21.Fw | gagctcgaattcatcgatgaTAATTCTACGTGCGTTTTCTTCC | |
| A179 | Amp.intg.3'gga21.Rv | TTTTGGTTCGAGATTCTCCC | |
| A210 | Mut.Set2.C261A.Rv | ATTAGGTCTGGCAGAGTGATTGCAAAATCTTGCG | |
| A211 | Mut.Set2.C261A.Fw | AATCACTCTGCCAGACCTAATTGTTATGTTGATAAATGG | |

### A.7.7 Extraction of genomic DNA from yeast

Freshly growing yeast cells were collected using blue loop and moved to 2 ml screw-cap tube filled with 1 ml H20. Yeast cells were pelleted by spinning at 3600 rpm for 1 min before adding ~200 µl zirconium beads (Biospec), 200 µl breaking buffer (2% triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH8, 1mM EDTA) and 200 µl phenol:chloroform:isoamyl alcohol (25:24:1 ; Sigma Aldrich). Yeast cells were homogenized by bead beating using Precellys 24 homogenizer (Bertin) at 5000 rpm for two cycles, 20 sec each. Tubes were centrifuged at 12000 rpm for 6 min, and the upper aqueous layer was moved to a fresh tube before adding an equal volume of chloroform and vortexing for 1 min. After centrifuging at maximum speed for 1 min, the aqueous layer was moved to a fresh tube and chloroform washing was repeated. The aqueous layer was moved to a fresh tube for 2.5 volumes of cold 100 EtOH to be added and mixed by inversion. The tubes were incubated at -20 °C for one hour then centrifuged at 14000 °C for 1 min to pellet the DNA. DNA pellet was washed with 500 µl cold 70% EtOH and centrifuged again. EtOH was completely removed and pellet was air dried before suspending in 30 µl TE buffer.

### A.7.8 Extraction of total RNA from yeast

Fixed number of yeast cells equivalent to 8 ODs were harvested from freshly growing liquid culture ($OD_{600}$ of 0.5 – 0.8) by centrifugation at 2600 rpm for 2 min at room temperature. Yeast cells pellet was resuspended in nuclease free water and moved to a fresh tube to be pelleted again. After discarding supernatant, 1 ml of TriReagent (Sigma Aldrich) was added to the pellet to be resuspended and moved to a new tube filled with 700 µl zirconium beads. Yeast cells were homogenized using Precellys at 5500 rpm for two cycles, 20 sec each. The tubes were pierced with hot needles and inserted in 5 ml glass tubes to be centrifuged at 1000 rpm for 1 min to bring lysate down. The lysate was resuspended, moved to a fresh tube and left to stand at room temperature for 5 min before centrifuging at maximum speed for 10 min at room temperature. The RNA-containing supernatant was moved to a fresh tube for 100 µl bromochloropropane to be added and vortexed for 15 sec then left to stand at room temperature for 10 min before centrifuging for 10 min at 4 °C. The upper aqueous layer was moved to a new tube to repeat the bromochloropropane washing step and for the upper aqueous layer to be aspirated and mixed with 500 µl isopropanol then left to stand at room temperature for 5 min before centrifuging for 10 min. to pellet the RNA. The RNA pellet was washed with 500

µl cold 75% EtOH then centrifuged for 5 min. The EtOH was completely removed and RNA pellet was air dried for 5 min before dissolving in 20 µl nuclease free water.

### A.7.9 Yeast cell lysis and preparation for dry and wet Western blotting

Yeast cell lysis and semi-dry Western blotting steps were carried out as indicated in Chapter 1 for the main materials and methods part of the thesis. For wet Western blotting, SDS-PAGE protein separation was performed using NuPAGE gel in Invitrogen mini gel tanks containing 1X MOPS buffer. For the blotting, nitrocellulose membrane was used. Antibodies used are listed in Table S2. All primary antibodies were used at 1:1000 dilution. Secondary antibodies were used at 1:2000 dilution.

### A.7.10 Reverse transcription and qPCR

Before reverse transcription, genomic DNA was eliminated from RNA using TURBO DNA-free$^{TM}$ Kit (Ambion, AM1907) following rigorous treatment conditions. cDNA was synthesized from 0.5 – 1 µg total RNA using NEB ProtoScript II First Strand cDNA Synthesis Kit following manufacturer manual. In brief, RNA was denatured with 2 µl of 50 µM oligo d(T)23 VN or 60 µM random primer mix in 8 µl reaction volume by incubating the whole mixture at 65°C for 5 min then snap cooling on ice before adding 10 µl of 2X ProtoScript II reaction mix and 2 µl of 2X ProtoScript II Enzyme Mix and incubating at 42 °C for one hour. The reaction was inactivated by incubating at 80 °C for 5 min. No-RT negative control reaction was included.

DNA was serially diluted and triplicates were used for each dilution. qPCR reaction was prepared using NEB Luna Universal qPCR Master Mix in 20 µl reaction volume containing 10 µl of 2X master mix, 250 nM of each primer and 5 µl of diluted cDNA. qPCR reaction was run on 7900HT (Applied Biosciences) real time instrument with program set for initial denaturing at 95 °C for 1 min, denaturing at 95 °C for 15 seconds and extending at 60 °C for 30 °C. Primers used for qPCRs are listed in Table A.5. Melting curve was generated using instrument recommended conditions with 60 °C for annealing and 95 °C for denaturing.

Data from Ct values was analyzed using ΔΔCt method. Average Ct values for target genes were normalized to those of gpd3 for each experimental sample before calculating the fold difference.

**Table A.5. List of primers used for RT-PCR, RT-qPCR and CRIP-RT-qPCR experiments.**

| No | Name | Sequence | Target sequence/remarks |
|---|---|---|---|
| A136 | Fw.qPCR.H3.FLAG | CAGTTGGCTCGTCGTCTC | Binds to 3' end of H3 |
| A137 | Rv.qPCR.H3.FLAG | ACTTGTCATCGTCATCCTTGT | Reverse binds to unique FLAG sequence. |
| A142 | Fw.gpd3p.qPCR | TTCCATCCTTGCCTCTCTCG | Gpd3 promoter |
| A143 | Rv.gpd3p.qPCR | GGCGATTTCCGTGTTCGGTA | |
| A144 | Fw.cyc1p.qPCR | ACATTAGGTCCTTTGTAGCATAA | TATA*cyc1p* |
| A145 | Rv.cyc1p.qPCR | ACCTCCAGGATGATAAACGG | |
| A147 | Fw1.cyc1p-H3.1.Ch.TSS | atccgtttatcatcctggag | RT-PCR to check RNA upstream ATG and downstream TSS of TATA*cyc1p* |
| A148 | Fw2.cyc1p-H3.1.Ch.TSS | ACTAAATTACCGGATCAATTCG | |
| A149 | Fw3.cyc1p-H3.1.Ch.TSS | ACTTCTATAGACACGCAAACAC | |
| A150 | Fw4.cyc1p-H3.1.Ch.TSS | TAGGTCCTTTGTAGCATAAATTAC | |
| A151 | Fw5.cyc1p-H3.1.Ch.TSS | CTTATACATTAGGTCCTTTGTAGC | |
| A262 | qPCR.3LacZ.Fw | CCAGTTGGTCTGGTGTCAAA | Binds to 3' end of lacZ |
| Y190 | GAPDH.Fw | AACATCATCCCCTCCTCCAC | For GAPDH gene body |
| Y191 | GAPDH.Rv | GCCTTGATGTCCTCGTAGTTG | |

For reverse transcription followed by standard PCR and agarose gel electrophoresis, *Taq* PCR was conducted using the same target specific primers and concentrations, but with minimal cycling (28 to 30 cycles) to check signal strength after running on agarose gel. The same strategy was adopted to check and compare transcription start site and upstream end of transcripts.

### A.7.11 5' RACE and TSS check primers

5' RACE strategy was adopted from and designed similar to Invitrogen Life technologies manual for 5′ RACE System for Rapid Amplification of cDNA Ends, Version 2.0. Primers used for 5' RACE are listed in Table A.6. The strategy depends on 3' end tailing of gene-specific cDNA followed by primary PCR using a reverse nested primer and forward primer that introduces a linker sequence and generates cDNA second strand for a second nested PCR that introduces sequences for prospective restriction digestion sites. Final PCR product was either run on agarose gel for preliminary band detection and fragment size verification, or digested and cloned for sequencing. After RNA extraction and DNase treatment, cDNA was synthesized from 1 µg of RNA using the same kit indicated above with lacZ gene specific primer GSP1. The resulting cDNA was purified using Promega PCR and gel clean up kit before the 3' poly-dC tailing reaction. The tailing reaction was done in 25 µl reaction volume containing purified cDNA, 1X TdT buffer, 1X CoCl2, 5 pmoles dCTPs, and 10 U TdT (NEB)

in a thermocycler with program set at 37 °C for 45 min for tailing followed by 70 °C for 15 min for inactivation.

1 µl of the tailing reaction was used directly as a template for 20 µl PCR reaction using Promega *Taq* PCR and following manufacturer manual. PCR product was diluted at 1:50 dilution before the second nested PCR using the same conditions. For cloning 5' RACE products the nested PCR was done using Primers with restriction sites, and the PCR product was purified then double digested in a single restriction reaction containing 1X Cutsmart buffer, 10 units of each enzyme in 50 µl reaction volume for 1 hour at 37 °C before purifying and ligating to a plasmid vector.

**Table A.6. List of primers used for lacZ 5' RACE experiment.**

| No. | Name | Sequence | Application |
|---|---|---|---|
| A229 | 5'RACE.AAP | GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG | Extension of poly-dC tailed cDNA |
| A230 | 5'RACE.AUAP | GGCCACGCGTCGACTAGTAC | Forward linker primer for primary PCR |
| A231 | LacZ.GSP1 | GCTCAGGTCAAATTCAGACG | LacZ first strand cDNA synthesis |
| A232 | LacZ.GSP2 | TTGCACCACAGATGAAACGC | Antisense nested primer for primary PCR |
| A233 | LacZ.GSP3 | TAACAACCCGTCGGATTCTC | Antisense nested primer for second PCR |
| A241 | NcoI-AUAP-dT(20)VN | ttacATCCGGTCGACTAGTACTTTTTTTTTTTTTTTTTTTTVN | Oligo-dT primer with linker |
| A242 | LacZ.GSP3.XmaI | ttaacccgggTAACAACCCGTCGGATTCTC | Antisense nested primer for second PCR and cloning |
| A248 | NcoI-AUAP | ttacCCATGGTCGACTAGTAC | |

### A.7.12 Capped RNA IP and qPCR (CRIP-qPCR)

Immunoprecipitation of capped RNA was performed as previously described with minor modifications **(Jimeno-Gonzalez *et al.*, 2010)**. RNA was extracted from 100 ml overnight freshly growing yeast culture with or without adding 10 µM doxycycline. DNA was eliminated using the TURBO DNase free kit as described above. 1 µg of RNA was used for the input and for each IP sample. 50 µl of protein A/protein G mix was used for each sample. Beads were

first washed three times, each with 1 ml of IPP150 buffer (150 mM NaCl, 0.1% NP-40, 10 mM Tris, pH 8.0) then washed similarly three time with 1 ml of IPP500 buffer (500 mM NaCl, 0.1% NP-40, 10 mM Tris, pH 8.0). 50 µl IPP500 buffer containing 10 µg of anti m7G-ca-RN106M (MBL) antibody was added to the beads and incubated for 2 hours at 4 °C with rotation. No-antibody control was included. Beads were washed three times with 1 ml of IPP150 buffer then added to the immunoprecipitation mix ((2.5 µg of yeast RNA, 5 µl of 0.1M DTT, 50U of RiboLock RNase inhibitor (ThermoFischer Scientific) and IPP150 buffer to 200 µl total volume). Immunoprecipitation mix was rotated overnight at 4°C. Beads were washed five times, each with 0.5 ml of IPP150 buffer/2.5 mM DTT, then resuspended in 200 µl of IPP150 buffer with 1mg/ml Proteinase K (Ambion) to be incubated at 37 °C for 30 min. Phenol/chloroform extraction and ethanol precipitation was performed as above, the RNA pellets were resuspended in 20 µl nuclease free water and diluted for cDNA synthesis. cDNA synthesis was performed using ProtoScript II kit (NEB) and hexamer primers. No-RT control was included for all samples. TATA*cyc1p* and gpd3 (GAPDH) control gene used for normalization were amplified using primers in Table A.5. Standard curve was generated using log concentration vs average Ct values for each dilution triplicates. Expected values were generated using interpolation.

# Supplementary

**Table S1. List of *S. pombe* strains used for R-loop project work.**

| Strain No. | Strain title | Genotype | Reference |
|---|---|---|---|
| P20 | WT | *h+, leu1-32, ura4 DS/E, ade6-216, HIS+, Otr1R::Ura4 (SphI)* | Lab strain |
| P49 | *rnh1Δ::NatNT2 & rnh201Δ::hphNT1 /1* | *h+, leu1-32, ura4 DS/E, ade6-210, HIS+, Otr1R::Ura4 (SphI), rnh1Δ::NatNT2, rnh201Δ::hphNT1,* | Lab strain |
| P50 | *rnh1Δ::NatNT2 & rnh201Δ::hphNT1 /2* | *Mat1Msmt0, leu1-32, ura4 DS/E, ade6-210, his2, Otr1R::Ura4 (SphI), rnh1Δ::NatNT2, rnh201Δ::hphNT1* | Lab strain |

**Table S2. List of antibodies used in this study**

| Factor | Antibody supplier & cat. # |
|---|---|
| AhR | ThermoFischer, MA1-513 |
| Flag epitope | Sigma, clone M2, A8592 |
| H3K36me3 | Abcam, Ab9070 |
| H3K9me2 | Abcam, Ab1220 |
| Hexokinase | Novus biologicals, NB120-20547 |
| M7G | MBL, RN106M |
| PAP | Sigma, P1291 |
| R-loop | Millipore, S9.6, MABE1095 |
| S2 | Millipore, clone 3E10, 04-1571 |
| S5 | Millipore, clone 3E8, 04-1572 |
| S7 | Millipore, clone 4E12, 04-1570 |
| Thr4 | Active motif, clone 6D7, 61361 |
| Y1 | Millipore, clone 3D12, MABE350 |

**Table S3. List and components of enzymatic reactions for on-bead splint then in-solution A-T ligation**

| **ChIP-exo-SL-DS** | | | |
|---|---|---|---|
| **No** | **Reaction** | **Components** | **Volume (µl)** |
| **1** | Polishing/blunting | Beads with the chromatin (20 µl), 10mM Tris-Cl, pH 8.0 (27 µl), 10X NEB buffer 2 (6 µl) [1X], 10X BSA (3 µl) [0.5X], 3mM dNTPs (3 µl) [150 µM each], 3 U/µl T4 DNA polymerase (1 µl) [3 U]. | 60 |
| **2** | First kinase | Beads with the chromatin (20 µl), 10mM Tris-Cl, pH 7.5 (33 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 10 U/µl T4 polynucleotide kinase (1 µl) [10 U]. | 60 |
| **3** | lambda exonuclease | Beads with the chromatin (20 µl), 10 mM Tris-Cl, pH 9.2 (32 µl), 10X λ exo buffer (6 µl) [1X], 5 U/µl λ exonuclease (4 µl) [20 U] | 60 |
| **4** | $Rec_{Jf}$ exonuclease | Beads with the chromatin (20 µl), 10 mM Tris-Cl, pH 8 (33 µl), 10X NEB buffer 2 (6 µl) [1X], 30 U/µl $Rec_{Jf}$ exonuclease (2.5 µl) [75 U] | 60 |

| | | | |
|---|---|---|---|
| **5** | First adaptor ligation | Beads with the chromatin (20 µl), 10 mM Tris-Cl, pH 7.5 (28 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 5 µM Nl5Aa-SL splint adaptor (4.5 µl) [0.375 µM], 400 U/µl T4 DNA ligase (3 µl) [1200 U] | 60 |
| **Elution, RNA digestion, decrosslinking, and DNA extraction and precipitation** | | | |
| **8** | Denaturing, annealing | DNA (11 µl), 10X Phi 29 DNA Pol. buffer (2 µl) [1X], 10X BSA (4 µl) [2X], 3 mM dNTPs (1 µl) [150 µM], 20 µM Extension primer (1 µl) [1 µM]. | 19 |
| **9** | Primer extension | DNA (11 µl), 10X Phi 29 DNA Pol. buffer (2 µl) [1X], 10X BSA (4 µl) [2X], 3 mM dNTPs (1 µl) [150 µM], 20 µM extension primer (1 µl) [1 µM], 10 U/µl Phi 29 DNA Pol. (1 µl) [10 U] | 20 |
| **10** | Second dA-tailing | DNA (20 µl), 1X TE (4.5 µl), 10X *Taq* buffer (3 µl) [1X], 3 mM dATPs (2 µl) [200 µM], 5 U/µl Taq Pol. (0.5 µl) [2.5 U] | 30 |
| **AMPure XP beads purification** | | | |
| **11** | Second ligation | DNA (30 µl), 1X TE (4 µl), 10X T4 DNA ligase buffer (4 µl) [1X], 15 uM A-T NL5Aa (1 µl) [0.375 uM], 400 U/µl T4 DNA ligase (1.25 µl) [500 U], | 40 |

**Table S4. List and components of enzymatic reactions for on-bead A-T the in-solution splint ligation**

| **ChIP-exo-DS-SL** | | | |
|---|---|---|---|
| **No** | **Reaction** | **Components** | **Volume (µl)** |
| **1** | Polishing/ blunting | Beads with the chromatin (20 µl), 10mM Tris-Cl, pH 8.0 (27 µl), 10X NEB buffer 2 (6 µl) [1X], 10X BSA (3 µl) [0.5X], 3mM dNTPs (3 µl) [150 µM each], 3 U/µl T4 DNA polymerase (1 µl) [3 U]. | 60 |
| **2** | First kinase | Beads with the chromatin (20 µl), 10mM Tris-Cl, pH 7.5 (33 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 10 U/µl T4 polynucleotide kinase (1 µl) [10 U]. | 60 |
| **3** | First dA-tailing | Beads with the chromatin (20 µl), 10 mM Tris-Cl, pH 8 (31 µl), 10X NEB buffer 2 (6 µl) [1X], 3 mM dATPs (2 µl) [100 µM], 5 U/µl Klenow exo- (1 µl) [5 U] | 60 |
| **4** | First adaptor ligation | Beads with the chromatin (20 µl), 10 mM Tris-Cl, pH 7.5 (28 µl), 10S T4 DNA ligase buffer (6 µl) [1X], 15 µM small A-T adaptor (5 µl) [1.25 µM], 400 U/µl T4 DNA ligase (1.25 µl) [500 U] | 60 |
| **5** | Second Kinase | Beads with the chromatin (20 µl), 10 mM Tris-Cl, pH 7.5 (33 µl), 10X T4 DNA ligase buffer (6 µl) [1X], 10 U/µl T4 PNK (1 µl) [10 U]. | 60 |
| **6** | lambda exonuclease | Beads with the chromatin (20 µl), 10 mM Tris-Cl, pH 9.2 (32 µl), 10X λ exo buffer (6 µl) [1X], 5 U/µl λ exonuclease (3 µl) [15 U] | 60 |
| **7** | $Rec_{Jf}$ exonuclease | Beads with the chromatin (20 µl), 10 mM Tris-Cl, pH 8 (33 µl), 10X NEB buffer 2 (6 µl) [1X], 30 U/µl $Rec_{Jf}$ exonuclease (1.5 µl) [45 U] | 60 |
| **Elution, RNA digestion, decrosslinking, and DNA extraction and precipitation** | | | |
| **8** | Second splint ligation | DNA (30 µl), 1X TE (4 µl), 10X T4 DNA ligase buffer (4 µl) [1X], 10 uM Uni-SL (1.5 µl) [0.375 uM], 400 U/µl T4 DNA ligase (3 µl) [1200 U], | 40 |

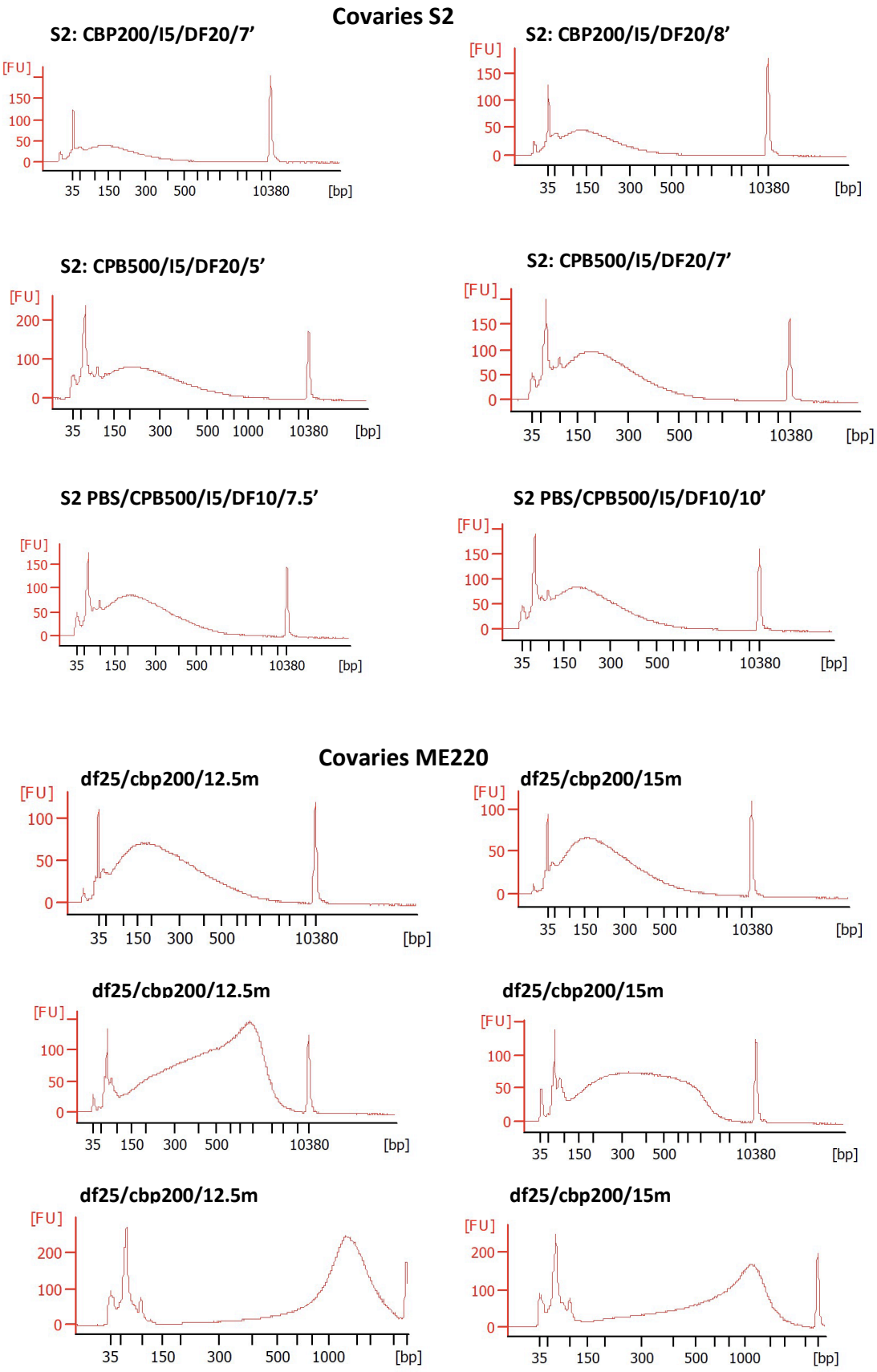



**Figure S1. DNA fragment size and sonication profiles using Covaries S2 &ME220 device.** Bioanalyzer profiles showing reproducibility of sonication using S2 device but not ME220 device. Sonication conditions are indicated.

**Table S5. Reactions components and conditions used for optimizing the SMART for R-loop mapping using DRIPc.**

| Step / Sample | TSAP or PNK + ATP | Purif. | Poly-A Polymerase | Purif. | Reverse transcription |
|---|---|---|---|---|---|
| 1 | No treatment | No purification | RNA, 1X FS buffer, 1 µl of 2 mM ATP, 20 U RNasin (Promega), 5 U PAP (NEB; M0276S) in 12 µl total volume. 37 °C/30 min, 65 °C/20 min | No purification | RNA, 1 µl of 10 mM dNTPs, 1 µl of 2µM dT24VN in 14 µl total volume. 65 °C/5 min, snap chilled then added; 1X FS buffer, 2 µl of 100 mM DTT, 20 U RNasin, 100 U SSII RT (ThermoFischer Scientific; 18064014) for a final volume of 20 µl. Later, added 1 µl of 10 µM TSP |
| 2 | No treatment | | No treatment | | |
| 3 | RNA, 1X FS buffer, 20 U RNasin (Promega), 1 U TSAP (Promega; M9910) in 10 µl total volume. 37 °C/15 min, 74 °C/15 min | | RNA, 1X FS buffer, 1 µl of 2 mM ATP, 20 U RNasin (Promega), 5 U PAP (NEB; M0276S) in 12 µl total volume. 37 °C/30 min, 65 °C/20 min | | |
| 4 | RNA, 1X T4 DNA ligase buffer, 20 U RNasin (Promega), 10 U PNK (NEB; M0201S) in 20 µl total volume. 37 °C/20 min, no heat inactivation | Zymoresearch 25 kit | RNA, 1X PAP buffer, 1 µl of 5 mM ATP, 20 U RNasin (Promega), 5 U PAP (NEB; M0276S) in 27 µl total volume. 37 °C/30 min, 65 °C/20 min | Zymoresearch 25 kit | The same steps and components of the RT reaction above, but with double the volume of the components:<br><br>RNA, 2 µl of 10 mM dNTPs, 2 µl of 2µM dT24VN in 26 µl total volume.<br><br>65 °C/5 min, snap chilled then added;<br><br>1X FS buffer, 4 µl of 100 mM DTT, 40 U RNasin, 200 U SSII RT (ThermoFischer Scientific; 18064014) for a final volume of 40 µl.<br><br>42 °C/2 min, room temp. Before adding RT enzyme,<br><br>42 °C/20 min, room temp<br><br>Added 2 µl of 10 µM TSP. |
| 5 | RNA, 1X Multicore buffer, 20 U RNasin (Promega), 1 U TSAP (Promega; M9910) in 20 µl total volume. 37 °C/20 min, no heat inactivation | | | | |
| 6 | RNA, 1X PAP buffer, 10 mM ATP, 20 U RNasin (Promega), 10 U PNK (NEB; M0201S) in 10 µl total volume. 37 °C/20 min, 74 °C/15 min | No purification | RNA, 1X PAP buffer, 1 µl of 5 mM ATP, 20 U RNasin (Promega), 5 U PAP (NEB; M0276S) in 20 µl total volume. 37 °C/30 min, 65 °C/20 min | Zymoresearch 25 kit | |
| 7 | RNA, 1X PAP buffer, 20 U RNasin (Promega), 1 U TSAP (Promega; M9910) in 10 µl total volume. 37 °C/20 min, 74 °C/15 min | | | | |

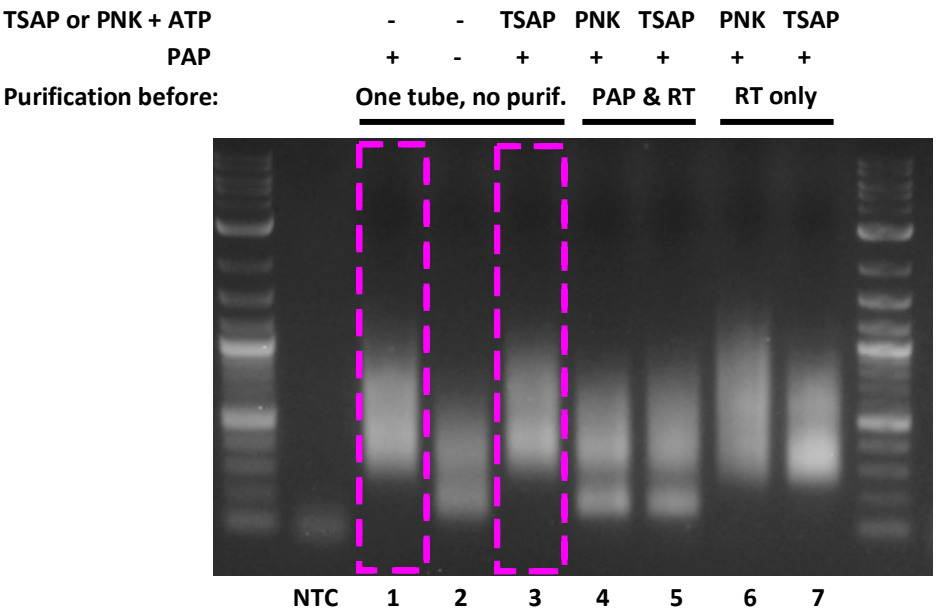



**Figure S5. Robust amplification for DRIPc RNA processed by SMART method.** All conditions used and steps are detailed in Table 4.2. For this, I treated the RNA with either thermostable alkaline phosphatase to dephosphorylate both ends of RNA or a polynucleotide kinase (PNK) plus ATPs to phosphorylate RNA 5' ends. Moreover, due to concerns about efficiency of the successive reactions due to carryover from the former ones, I included column purification steps at some stages in between the different reactions for some of the samples. I also included a no-tailing enzyme control to check the dependence on the poly-A tail. 2 µl of or equivalent were used from the SMART output for the check PCR amplification. RNA sample was treated with different modification enzymes (either PNK + ATP, TSAP, or none), poly-A tailed (except no-tailing control) and finally reverse transcribed. 2 µl of SMART product were used for PCR amplification with primers for 27 cycles following Taq program. Purification steps were included between some reactions for some samples. 20 µl were separated on 1% agarose gel. 1 µg 2-log ladder was run in parallel.

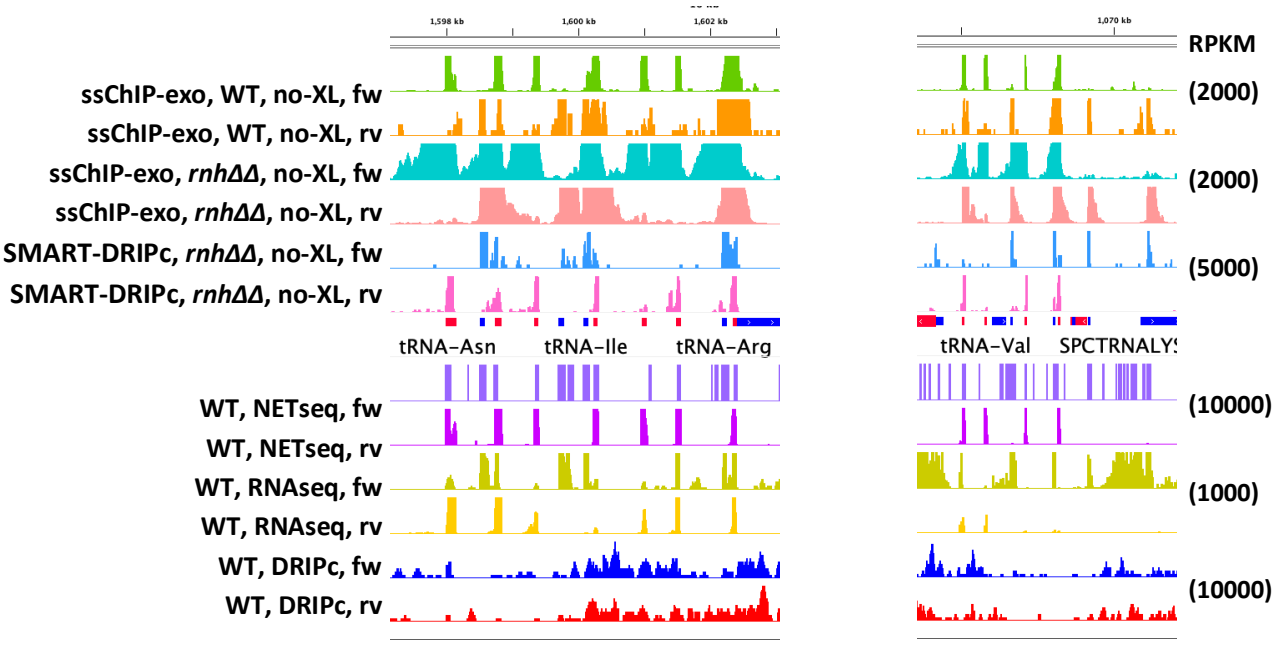



**Figure S9. R-loop signals over tRNAs don't overlap with dsRNA signals.**

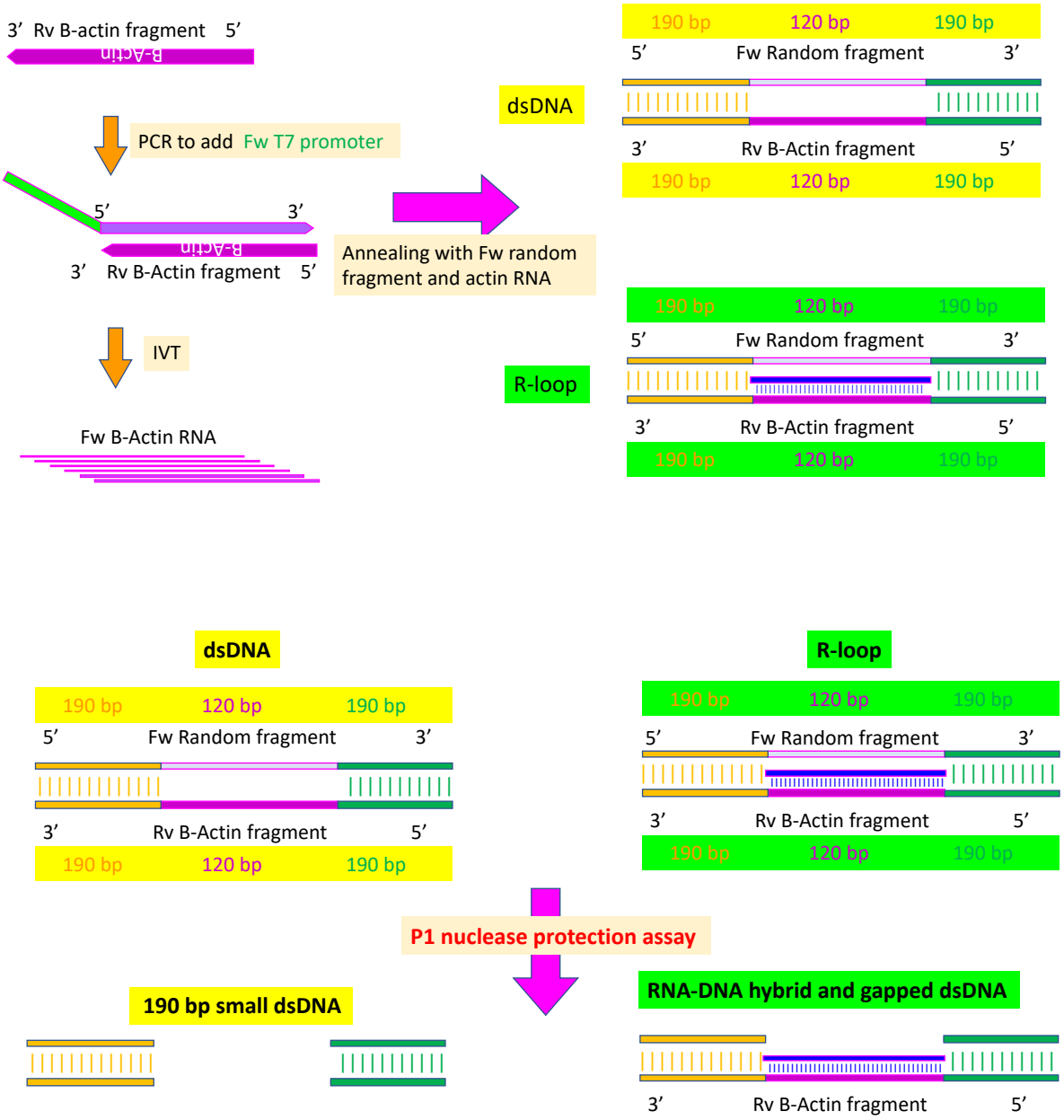



**Figure S13. Outline for artificial R-loop preparation strategy.**

# References


Agarwal, S., Macfarlan, T. S., Sartor, M. A., & Iwase, S. (2015). Sequencing of first-strand cDNA library reveals full-length transcriptomes. *Nat Commun, 6*, 6002. doi:10.1038/ncomms7002

Aguilera, A. (2002). The connection between transcription and genomic instability. *EMBO J, 21*(3), 195-201. doi:10.1093/emboj/21.3.195

Aguilera, A., & Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol Cell, 46*(2), 115-124. doi:10.1016/j.molcel.2012.04.009

Al-Hadid, Q., & Yang, Y. (2016). R-loop: an emerging regulator of chromatin dynamics. *Acta Biochim Biophys Sin (Shanghai), 48*(7), 623-631. doi:10.1093/abbs/gmw052

Alecki, C., Chiwara, V., Sanz, L. A., Grau, D., Arias Perez, O., Boulier, E. L., . . . Francis, N. J. (2020). RNA-DNA strand exchange by the Drosophila Polycomb complex PRC2. *Nat Commun, 11*(1), 1781. doi:10.1038/s41467-020-15609-x

Allison, D. F., & Wang, G. G. (2019). R-loops: formation, function, and relevance to cell stress. *Cell Stress, 3*(2), 38-46. doi:10.15698/cst2019.02.175

Allshire, R. C., & Ekwall, K. (2015). Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. *Cold Spring Harb Perspect Biol, 7*(7), a018770. doi:10.1101/cshperspect.a018770

Allshire, R. C., & Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol, 19*(4), 229-244. doi:10.1038/nrm.2017.119

Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schafer, A., Grummt, I., & Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nat Genet, 51*(2), 217-223. doi:10.1038/s41588-018-0306-6

Ariel, F., Lucero, L., Christ, A., Mammarella, M. F., Jegu, T., Veluchamy, A., . . . Crespi, M. (2020). R-Loop Mediated trans Action of the APOLO Long Noncoding RNA. *Mol Cell, 77*(5), 1055-1065 e1054. doi:10.1016/j.molcel.2019.12.015

Arnott, S., Chandrasekaran, R., Millane, R. P., & Park, H. S. (1986). DNA-RNA hybrid secondary structures. *J Mol Biol, 188*(4), 631-640. doi:10.1016/s0022-2836(86)80011-0

Atkinson, S. R., Marguerat, S., Bitton, D. A., Rodríguez-López, M., Rallis, C., Lemay, J.-F., . . . Bähler, J. (2018). Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA (New York, N.Y.), 24*(9), 1195-1213. doi:10.1261/rna.065524.118

Audergon, P. N., Catania, S., Kagansky, A., Tong, P., Shukla, M., Pidoux, A. L., & Allshire, R. C. (2015). Epigenetics. Restricted epigenetic inheritance of H3K9 methylation. *Science, 348*(6230), 132-135. doi:10.1126/science.1260638

Ausubel, F., Dracopoli, N., & Wildermuth, M. (2011). Current Protocols in Molecular Biology. In (Vol. 96).

Baaklini, I., Hraiky, C., Rallu, F., Tse-Dinh, Y. C., & Drolet, M. (2004). RNase HI overproduction is required for efficient full-length RNA synthesis in the absence of topoisomerase I in Escherichia coli. *Mol Microbiol, 54*(1), 198-211. doi:10.1111/j.1365-2958.2004.04258.x

Backert, S. (2002). R-loop-dependent rolling-circle replication and a new model for DNA concatemer resolution by mitochondrial plasmid mp1. *EMBO J, 21*(12), 3128-3136. doi:10.1093/emboj/cdf311

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K., & Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nat Struct Mol Biol, 20*(10), 1199-1205. doi:10.1038/nsmb.2662

Belmont, P., Constant, J. F., & Demeunynck, M. (2001). Nucleic acid conformation diversity: from structure to function and regulation. *Chemical Society Reviews, 30*(1), 70-81. doi:10.1039/a904630e

Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M., & Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. *Science, 301*(5636), 1090-1093. doi:10.1126/science.1085703

Belotserkovskii, B. P., De Silva, E., Tornaletti, S., Wang, G., Vasquez, K. M., & Hanawalt, P. C. (2007). A triplex-forming sequence from the human c-MYC promoter interferes with DNA transcription. *J Biol Chem, 282*(44), 32433-32441. doi:10.1074/jbc.M704618200

Belotserkovskii, B. P., & Hanawalt, P. C. (2011). Anchoring nascent RNA to the DNA template could interfere with transcription. *Biophys J, 100*(3), 675-684. doi:10.1016/j.bpj.2010.12.3709

Belotserkovskii, B. P., & Hanawalt, P. C. (2015). PNA binding to the non-template DNA strand interferes with transcription, suggesting a blockage mechanism mediated by R-loop formation. *Mol Carcinog, 54*(11), 1508-1512. doi:10.1002/mc.22209

Belotserkovskii, B. P., Liu, R., Tornaletti, S., Krasilnikova, M. M., Mirkin, S. M., & Hanawalt, P. C. (2010). Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc Natl Acad Sci U S A, 107*(29), 12816-12821. doi:10.1073/pnas.1007580107

Belotserkovskii, B. P., Mirkin, S. M., & Hanawalt, P. C. (2013). DNA sequences that interfere with transcription: implications for genome function and stability. *Chem Rev, 113*(11), 8620-8637. doi:10.1021/cr400078y

Belotserkovskii, B. P., Neil, A. J., Saleh, S. S., Shin, J. H., Mirkin, S. M., & Hanawalt, P. C. (2013). Transcription blockage by homopurine DNA sequences: role of sequence composition and single-strand breaks. *Nucleic Acids Res, 41*(3), 1817-1828. doi:10.1093/nar/gks1333

Belotserkovskii, B. P., Soo Shin, J. H., & 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 Res, 45*(11), 6589-6599. doi:10.1093/nar/gkx403

Belotserkovskii, B. P., Tornaletti, S., D'Souza, A. D., & Hanawalt, P. C. (2018). R-loop generation during transcription: Formation, processing and cellular outcomes. *DNA Repair (Amst), 71*, 69-81. doi:10.1016/j.dnarep.2018.08.009

Benore-Parsons, M., & Ayoub, M. A. (1997). Presence of RNase A causes aberrant DNA band shifts. *Biotechniques, 23*(1), 128-131. doi:10.2144/97231st04

Bernecky, C., Herzog, F., Baumeister, W., Plitzko, J. M., & Cramer, P. (2016). Structure of transcribing mammalian RNA polymerase II. *Nature, 529*(7587), 551-554. doi:10.1038/nature16482

Bettin, N., Oss Pegorar, C., & Cusanelli, E. (2019). The Emerging Roles of TERRA in Telomere Maintenance and Genome Stability. *Cells, 8*(3). doi:10.3390/cells8030246

Bhatia, V., Barroso, S. I., Garcia-Rubio, M. L., Tumini, E., Herrera-Moyano, E., & Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature, 511*(7509), 362-365. doi:10.1038/nature13374

Bhattacharyya, A., Murchie, A. I., & Lilley, D. M. (1990). RNA bulges and the helical periodicity of double-stranded RNA. *Nature, 343*(6257), 484-487. doi:10.1038/343484a0

Birnstiel, M. L., Sells, B. H., & Purdom, I. F. (1972). Kinetic complexity of RNA molecules. *J Mol Biol, 63*(1), 21-39. doi:10.1016/0022-2836(72)90519-0

Blake, R. D., & Delcourt, S. G. (1996). Thermodynamic Effects of Formamide on DNA Stability. *Nucleic Acids Research, 24*(11), 2095-2103. doi:10.1093/nar/24.11.2095

Bochman, M. L., Paeschke, K., & Zakian, V. A. (2012). DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet, 13*(11), 770-780. doi:10.1038/nrg3296

Boguslawski, S. J., Smith, D. E., Michalak, M. A., Mickelson, K. E., Yehle, C. O., Patterson, W. L., & Carrico, R. J. (1986). Characterization of monoclonal antibody to DNA · RNA and

its application to immunodetection of hybrids. *J Immun Meth, 89*(1), 123-130. doi:10.1016/0022-1759(86)90040-2

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120. doi:10.1093/bioinformatics/btu170

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., . . . Guil, S. (2015). Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proc Natl Acad Sci U S A, 112*(18), 5785-5790. doi:10.1073/pnas.1421197112

Bronner, C., Salvi, L., Zocco, M., Ugolini, I., & Halic, M. (2017). Accumulation of RNA on chromatin disrupts heterochromatic silencing. *Genome Res, 27*(7), 1174-1183. doi:10.1101/gr.216986.116

Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., & Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res, 34*(19), 5402-5415. doi:10.1093/nar/gkl655

Cam, H. P., & Whitehall, S. (2016). Analysis of Heterochromatin in Schizosaccharomyces pombe. *Cold Spring Harb Protoc, 2016*(11). doi:10.1101/pdb.top079889

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., . . . Workman, J. L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell, 123*(4), 581-592. doi:10.1016/j.cell.2005.10.023

Castel, S. E., Ren, J., Bhattacharjee, S., Chang, A. Y., Sanchez, M., Valbuena, A., . . . Martienssen, R. A. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. *Cell, 159*(3), 572-583. doi:10.1016/j.cell.2014.09.031

Cerritelli, S. M., & Crouch, R. J. (2009). Ribonuclease H: the enzymes in eukaryotes. *FEBS J, 276*(6), 1494-1505. doi:10.1111/j.1742-4658.2009.06908.x

Chalamcharla, V. R., Folco, H. D., Dhakshnamoorthy, J., & Grewal, S. I. (2015). Conserved factor Dhp1/Rat1/Xrn2 triggers premature transcription termination and nucleates heterochromatin to promote gene silencing. *Proc Natl Acad Sci U S A, 112*(51), 15548-15555. doi:10.1073/pnas.1522127112

Chan, Y. A., Aristizabal, M. J., Lu, P. Y., Luo, Z., Hamza, A., Kobor, M. S., . . . Hieter, P. (2014). Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet, 10*(4), e1004288. doi:10.1371/journal.pgen.1004288

Chávez, S., Beilharz, T., Rondón, A. G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J. Q., . . . Aguilera, A. (2000). A protein complex containing Tho2, Hpr1, Mft1 and a novel

protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *EMBO J, 19*(21), 5824-5834. doi:10.1093/emboj/19.21.5824

Chedin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. *Trends Genet, 32*(12), 828-838. doi:10.1016/j.tig.2016.10.002

Chedin, F., & Benham, C. J. (2020). Emerging roles for R-loop structures in the management of topological stress. *J Biol Chem, 295*(14), 4684-4695. doi:10.1074/jbc.REV119.006364

Chédin, F., Hartono, S. R., Sanz, L. A., & Vanoosthuyse, V. (2021). Best practices for the visualization, mapping, and manipulation of R-loops. *EMBO J, 40*(4), e106394. doi:10.15252/embj.2020106394

Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W., & Tyler, J. K. (2015). The Overlooked Fact: Fundamental Need for Spike-In Control for Virtually All Genome-Wide Analyses. *Mol Cell Biol, 36*(5), 662-667. doi:10.1128/MCB.00970-14

Chen, L., Chen, J. Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., . . . Fu, X. D. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol Cell, 68*(4), 745-757 e745. doi:10.1016/j.molcel.2017.10.008

Cheng, Z., Otto, G. M., Powers, E. N., Keskin, A., Mertins, P., Carr, S. A., . . . Brar, G. A. (2018). Pervasive, Coordinated Protein-Level Changes Driven by Transcript Isoform Switching during Meiosis. *Cell, 172*(5), 910-923 e916. doi:10.1016/j.cell.2018.01.035

Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R., & Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. *PLoS Biol, 6*(11), e277. doi:10.1371/journal.pbio.0060277

Choi, J., & Majima, T. (2011). Conformational changes of non-B DNA. *Chemical Society Reviews, 40*(12), 5893-5909. doi:10.1039/C1CS15153C

Chou, S. H., Flynn, P., & Reid, B. (1989). High-resolution NMR study of a synthetic DNA-RNA hybrid dodecamer containing the consensus pribnow promoter sequence: d(CGTTATAATGCG).r(CGCAUUAUAACG). *Biochemistry, 28*(6), 2435-2443. doi:10.1021/bi00432a014

Cohen, S., Puget, N., Lin, Y. L., Clouaire, T., Aguirrebengoa, M., Rocher, V., . . . Legube, G. (2018). Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nat Commun, 9*(1), 533. doi:10.1038/s41467-018-02894-w

Costantino, L., & Koshland, D. (2015). The Yin and Yang of R-loop biology. *Curr Opin Cell Biol, 34*, 39-45. doi:10.1016/j.ceb.2015.04.008

Costantino, L., & Koshland, D. (2018). Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. *Mol Cell, 71*(4), 487-497 e483. doi:10.1016/j.molcel.2018.06.037

Cotobal, C., Rodriguez-Lopez, M., Duncan, C., Hasan, A., Yamashita, A., Yamamoto, M., . . . Mata, J. (2015). Role of Ccr4-Not complex in heterochromatin formation at meiotic genes and subtelomeres in fission yeast. *Epigenetics Chromatin, 8*, 28. doi:10.1186/s13072-015-0018-4

Cristini, A., Groh, M., Kristiansen, M. S., & Gromak, N. (2018). RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep, 23*(6), 1891-1905. doi:10.1016/j.celrep.2018.04.025

Crossley, M. P., Bocek, M., & Cimprich, K. A. (2019). R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell, 73*(3), 398-411. doi:10.1016/j.molcel.2019.01.024

Crossley, M. P., Bocek, M. J., Hamperl, S., Swigut, T., & Cimprich, K. A. (2020). qDRIP: a method to quantitatively assess RNA-DNA hybrid formation genome-wide. *Nucleic Acids Res, 48*(14), e84. doi:10.1093/nar/gkaa500

Daniels, G. A., & 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 Res, 23*(24), 5006-5011. doi:10.1093/nar/23.24.5006

de Jonge, W. J., O'Duibhir, E., Lijnzaad, P., van Leenen, D., Groot Koerkamp, M. J., Kemmeren, P., & Holstege, F. C. (2017). Molecular mechanisms that distinguish TFIID housekeeping from regulatable SAGA promoters. *EMBO J, 36*(3), 274-290. doi:10.15252/embj.201695621

Di, L., Fu, Y., Sun, Y., Li, J., Liu, L., Yao, J., . . . Wang, J. (2020). RNA sequencing by direct tagmentation of RNA/DNA hybrids. *Proc Natl Acad Sci U S A, 117*(6), 2886-2893. doi:10.1073/pnas.1919800117

Dona, F., & Houseley, J. (2014). Unexpected DNA loss mediated by the DNA binding activity of ribonuclease A. *PLoS One, 9*(12), e115008. doi:10.1371/journal.pone.0115008

Drolet, M., Bi, X., & Lius, L. F. (1994). Hypernegative Supercoiling of the DNA Template during Transcription Elongation in Vitro. *J Biol Chem, 269* ((3)), 2068-2074.

Drolet, M., Phoenix, P., Menzel, R., Masse, E., Liu, L. F., & Crouchs, R. J. (1995). Overexpression of RNase H partially complements the growth defect of an Escherichia coli AtopA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc. Natl. Acad. Sci., 92*, 3526-3530.

Drouin, S., Laramée, L., Jacques, P., Forest, A., Bergeron, M., & Robert, F. (2010). DSIF and RNA polymerase II CTD phosphorylation coordinate the recruitment of Rpd3S to actively transcribed genes. *PLoS Genet, 6*(10), e1001173. doi:10.1371/journal.pgen.1001173

Dumelie, J. G., & Jaffrey, S. R. (2017). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife, 6*. doi:10.7554/eLife.28306

Duncan, C. D., & Mata, J. (2014). The translational landscape of fission-yeast meiosis and sporulation. *Nat Struct Mol Biol, 21*(7), 641-647. doi:10.1038/nsmb.2843

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev, 18*(13), 1618-1629. doi:10.1101/gad.1200804

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M. E., & Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in E. coli. *Cell, 146*(4), 533-543. doi:10.1016/j.cell.2011.07.034

Egli, M., Usman, N., Zhang, S. G., & Rich, A. (1992). Crystal structure of an Okazaki fragment at 2-A resolution. *Proc Natl Acad Sci U S A, 89*(2), 534-538. doi:10.1073/pnas.89.2.534

El Hage, A., French, S. L., Beyer, A. L., & Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev, 24*(14), 1546-1558. doi:10.1101/gad.573310

El Hage, A., Webb, S., Kerr, A., & Tollervey, D. (2014). Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet, 10*(10), e1004716. doi:10.1371/journal.pgen.1004716

Evans, K. J., Huang, N., Stempor, P., Chesney, M. A., Down, T. A., & Ahringer, J. (2016). Stable Caenorhabditis elegans chromatin domains separate broadly expressed and developmentally regulated genes. *Proc Natl Acad Sci U S A, 113*(45), E7020-E7029. doi:10.1073/pnas.1608162113

Farnung, L., Vos, S. M., & Cramer, P. (2018). doi:10.1101/437574

Fedoroff, O., Salazar, M., & Reid, B. R. (1993). Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. *J Mol Biol, 233*(3), 509-523. doi:10.1006/jmbi.1993.1528

Feil, R., Sein, H., Värv, S., & Kristjuhan, A. (2015). Distribution and Maintenance of Histone H3 Lysine 36 Trimethylation in Transcribed Locus. *PLoS One, 10*(3). doi:10.1371/journal.pone.0120200

Ferrari, P., & Strubin, M. (2015). Uncoupling histone turnover from transcription-associated histone H3 modifications. *Nucleic Acids Res, 43*(8), 3972-3985. doi:10.1093/nar/gkv282

Fischer, T., Strässer, K., Rácz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., . . . Hurt, E. (2002). The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. *EMBO J, 21*(21), 5843-5852. doi:10.1093/emboj/cdf590

Fish, R. N., Bostick, M., Lehman, A., & Farmer, A. (2016). Transcriptome Analysis at the Single-Cell Level Using SMART Technology. *Curr Protoc Mol Biol, 116*, 4 26 21-24 26 24. doi:10.1002/cpmb.23

Formosa, T. (2008). FACT and the reorganized nucleosome. *Mol Biosyst, 4*(11), 1085-1093. doi:10.1039/b812136b

Formosa, T., Ruone, S., Adams, M. D., Olsen, A. E., Eriksson, P., Yu, Y., . . . Stillman, D. J. (2002). Defects in SPT16 or POB3 (yFACT) in Saccharomyces cerevisiae cause dependence on the Hir/Hpc pathway: polymerase passage may degrade chromatin structure. *Genetics, 162*(4), 1557-1571.

Fuchs, J., Dell'Atti, D., Buhot, A., Calemczuk, R., Mascini, M., & Livache, T. (2010). Effects of formamide on the thermal stability of DNA duplexes on biochips. *Anal Biochem, 397*(1), 132-134. doi:10.1016/j.ab.2009.09.044

Fujiwara, T., & Shindo, H. (1985). Phosphorus-31 nuclear magnetic resonance of highly oriented DNA fibers. 2. Molecular motions in hydrated DNA. *Biochemistry, 24*(4), 896-902. doi:10.1021/bi00325a013

Gansauge, M. T., Aximu-Petri, A., Nagel, S., & Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. *Nat Protoc, 15*(8), 2279-2300. doi:10.1038/s41596-020-0338-0

Garcia-Muse, T., & Aguilera, A. (2019). R Loops: From Physiological to Pathological Roles. *Cell, 179*(3), 604-618. doi:10.1016/j.cell.2019.08.055

Garcia-Pichardo, D., Canas, J. C., Garcia-Rubio, M. L., Gomez-Gonzalez, B., Rondon, A. G., & Aguilera, A. (2017). Histone Mutants Separate R Loop Formation from Genome Instability Induction. *Mol Cell, 66*(5), 597-609 e595. doi:10.1016/j.molcel.2017.05.014

Garcia-Rubio, M. L., Perez-Calero, C., Barroso, S. I., Tumini, E., Herrera-Moyano, E., Rosado, I. V., & Aguilera, A. (2015). The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLoS Genet, 11*(11), e1005674. doi:10.1371/journal.pgen.1005674

Gavalda, S., Gallardo, M., Luna, R., & Aguilera, A. (2013). R-loop mediated transcription-associated recombination in trf4Delta mutants reveals new links between RNA

surveillance and genome integrity. *PLoS One, 8*(6), e65541. doi:10.1371/journal.pone.0065541

Ginno, P. A., Lim, Y. W., Lott, P. L., Korf, I., & 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*(10), 1590-1600. doi:10.1101/gr.158436.113

Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., & Chedin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell, 45*(6), 814-825. doi:10.1016/j.molcel.2012.01.017

Glover, D. M., & Hogness, D. S. (1977). A novel arrangement of the 18S and 28S sequences in a repeating unit of drosophila melanogaster rDNA. *Cell, 10*, 167-176.

Gómez-González, B., & Aguilera, A. (2019). Transcription-mediated replication hindrance: a major driver of genome instability. *Genes Dev, 33*(15-16), 1008-1026. doi:10.1101/gad.324517.119

González-Aguilera, C., Tous, C., Gómez-González, B., Huertas, P., Luna, R., & Aguilera, A. (2008). The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol Biol Cell, 19*(10), 4310-4318. doi:10.1091/mbc.e08-04-0355

Govind, C. K., Qiu, H., Ginsburg, D. S., Ruan, C., Hofmeyer, K., Hu, C., . . . Hinnebusch, A. G. (2010). Phosphorylated Pol II CTD recruits multiple HDACs, including Rpd3C(S), for methylation-dependent deacetylation of ORF nucleosomes. *Mol Cell, 39*(2), 234-246. doi:10.1016/j.molcel.2010.07.003

Gowrishankar, J., & Harinarayanan, R. (2004). Why is transcription coupled to translation in bacteria? *Mol Microbiol, 54*(3), 598-603. doi:10.1111/j.1365-2958.2004.04289.x

Gowrishankar, J., Leela, J. K., & Anupama, K. (2013). R-loops in bacterial transcription: their causes and consequences. *Transcription, 4*(4), 153-157. doi:10.4161/trns.25101

Grabczyk, E., Mancuso, M., & Sammarco, M. C. (2007). A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic Acids Res, 35*(16), 5351-5359. doi:10.1093/nar/gkm589

Graf, M., Bonetti, D., Lockhart, A., Serhal, K., Kellner, V., Maicher, A., . . . Luke, B. (2017). Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. *Cell, 170*(1), 72-85 e14. doi:10.1016/j.cell.2017.06.006

Gray, D. M., & Ratliff, R. L. (1975). Circular dichroism spectra of poly[d(AC):d(GT)], poly[r(AC):r(GU)], and hybrids poly[d(AC):r(GU)] and poly[r(AC):d(GT)] in the presence of ethanol. *Biopolymers, 14*(3), 487-498. doi:10.1002/bip.1975.360140305

Grunseich, C., Wang, I. X., Watts, J. A., Burdick, J. T., Guber, R. D., Zhu, Z., . . . Cheung, V. G. (2018). Senataxin Mutation Reveals How R-Loops Promote Transcription by Blocking DNA Methylation at Gene Promoters. *Mol Cell, 69*(3), 426-437 e427. doi:10.1016/j.molcel.2017.12.030

Hacker, K. E., Fahey, C. C., Shinsky, S. A., Chiang, Y. J., DiFiore, J. V., Jha, D. K., . . . Rathmell, W. K. (2016). Structure/Function Analysis of Recurrent Mutations in SETD2 Protein Reveals a Critical and Conserved Role for a SET Domain Residue in Maintaining Protein Stability and Histone H3 Lys-36 Trimethylation. *J Biol Chem, 291*(40), 21283-21295. doi:10.1074/jbc.M116.739375

Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramskold, D., Hendriks, G. J., Larsson, A. J. M., . . . Sandberg, R. (2020). Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat Biotechnol, 38*(6), 708-714. doi:10.1038/s41587-020-0497-0

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol, 11*(5), 394-403. doi:10.1038/nsmb763

Hahn, S., Hoar, E. T., & Guarente, L. (1985). Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC-1 promoter of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A, 82*(24), 8562-8566. doi:10.1073/pnas.82.24.8562

Haist, K., Ziegler, C., & Botten, J. (2015). Strand-Specific Quantitative Reverse Transcription-Polymerase Chain Reaction Assay for Measurement of Arenavirus Genomic and Antigenomic RNAs. *PLoS One, 10*(5), e0120043. doi:10.1371/journal.pone.0120043

Halasz, L., Karanyi, Z., Boros-Olah, B., Kuik-Rozsa, T., Sipos, E., Nagy, E., . . . Szekvolgyi, L. (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res, 27*(6), 1063-1073. doi:10.1101/gr.219394.116

Hall, K. B., & McLaughlin, L. W. (1991). Thermodynamic and structural properties of pentamer DNA.DNA, RNA.RNA, and DNA.RNA duplexes of identical sequence. *Biochemistry, 30*(44), 10606-10613. doi:10.1021/bi00108a002

Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T., & Cimprich, K. A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell, 170*(4), 774-786 e719. doi:10.1016/j.cell.2017.07.043

Hamperl, S., & Cimprich, K. A. (2014). The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair, 19*, 84-94. doi:10.1016/j.dnarep.2014.03.023

Hartono, S. R., Malapert, A., Legros, P., Bernard, P., Chedin, F., & 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*(3), 272-284. doi:10.1016/j.jmb.2017.12.016

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., . . . Livingston, D. M. (2015). BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Mol Cell, 57*(4), 636-647. doi:10.1016/j.molcel.2015.01.011

Hazelbaker, D. Z., Marquardt, S., Wlotzka, W., & Buratowski, S. (2013). Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Mol Cell, 49*(1), 55-66. doi:10.1016/j.molcel.2012.10.014

He, Q., Johnston, J., & Zeitlinger, J. (2015). ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat Biotechnol, 33*(4), 395-401. doi:10.1038/nbt.3121

Hennig, B. P., Bendrin, K., Zhou, Y., & Fischer, T. (2012). Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. *EMBO Rep, 13*(11), 997-1003. doi:10.1038/embor.2012.146

Hennig, B. P., & Fischer, T. (2013). The great repression: chromatin and cryptic transcription. *Transcription, 4*(3), 97-101. doi:10.4161/trns.24884

Holder, I. T., Wagner, S., Xiong, P., Sinn, M., Frickey, T., Meyer, A., & Hartig, J. S. (2015). Intrastrand triplex DNA repeats in bacteria: a source of genomic instability. *Nucleic Acids Res, 43*(21), 10126-10142. doi:10.1093/nar/gkv1017

Holt, I. J. (2019). The mitochondrial R-loop. *Nucleic Acids Res, 47*(11), 5480-5489. doi:10.1093/nar/gkz277

Hu, Y., Bennett, H. W., Liu, N., Moravec, M., Williams, J. F., Azzalin, C. M., & King, M. C. (2019). RNA-DNA Hybrids Support Recombination-Based Telomere Maintenance in Fission Yeast. *Genetics, 213*(2), 431-447. doi:10.1534/genetics.119.302606

Huertas, P., & Aguilera, A. (2003). Cotranscriptionally Formed DNA:RNA Hybrids Mediate Transcription Elongation Impairment and Transcription-Associated Recombination. *Molecular Cell, 12*(3), 711-721. doi:10.1016/j.molcel.2003.08.010

Hung, S. H., Yu, Q., Gray, D. M., & Ratliff, R. L. (1994). Evidence from CD spectra that d(purine).r(pyrimidine) and r(purine).d(pyrimidine) hybrids are in different structural classes. *Nucleic Acids Res, 22*(20), 4326-4334. doi:10.1093/nar/22.20.4326

Huraiova, B., Kanovits, J., Polakova, S. B., Cipak, L., Benko, Z., Sevcovicova, A., . . . Gregan, J. (2020). Proteomic analysis of meiosis and characterization of novel short open reading frames in the fission yeast Schizosaccharomyces pombe. *Cell Cycle, 19*(14), 1777-1785. doi:10.1080/15384101.2020.1779470

Ito, S., Akamatsu, Y., Noma, A., Kimura, S., Miyauchi, K., Ikeuchi, Y., . . . Suzuki, T. (2014). A single acetylation of 18 S rRNA is essential for biogenesis of the small ribosomal

subunit in Saccharomyces cerevisiae. *J Biol Chem, 289*(38), 26201-26212. doi:10.1074/jbc.M114.593996

Jensen, T. H., Jacquier, A., & Libri, D. (2013). Dealing with pervasive transcription. *Mol Cell, 52*(4), 473-484. doi:10.1016/j.molcel.2013.10.032

Jiang, D., Zhang, X., Pang, Y., Zhang, J., Wang, J., & Huang, Y. (2019). Terminal transfer amplification and sequencing for high-efficiency and low-bias copy number profiling of fragmented DNA samples. *Protein Cell, 10*(3), 229-233. doi:10.1007/s13238-018-0540-9

Jiang, F., & Doudna, J. A. (2017). CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys. , 46*. doi:10.1146/annurev-biophys-

Jih, G., Iglesias, N., Currie, M. A., Bhanu, N. V., Paulo, J. A., Gygi, S. P., . . . Moazed, D. (2017). Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription. *Nature, 547*(7664), 463-467. doi:10.1038/nature23267

Jimeno-Gonzalez, S., Haaning, L. L., Malagon, F., & Jensen, T. H. (2010). The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II. *Mol Cell, 37*(4), 580-587. doi:10.1016/j.molcel.2010.01.019

Kaback, D. B., Angerer, L. M., & Davidson, N. (1979). Improved methods for the formation and stabilization of R-loops. *Nucleic Acids Res, 6*(7), 2499-2317. doi:10.1093/nar/6.7.2499

Kankia, B. I., & Marky, L. A. (1999). DNA, RNA, and DNA/RNA Oligomer Duplexes: A Comparative Study of Their Stability, Heat, Hydration, and Mg2+ Binding Properties. *The Journal of Physical Chemistry B, 103*(41), 8759-8767. doi:10.1021/jp991614x

Kasahara, M., Clikeman, J. A., Bates, D. B., & Kogoma, T. (2000). RecA protein-dependent R-loop formation in vitro. *Genes Dev, 14*, 360–365.

Katahira, J., Senokuchi, K., & Hieda, M. (2020). Human THO maintains the stability of repetitive DNA. *Genes Cells*. doi:10.1111/gtc.12760

Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2020). Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc, 15*(10), 3264-3283. doi:10.1038/s41596-020-0373-x

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., . . . Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun, 10*(1), 1930. doi:10.1038/s41467-019-09982-5

Keogh, M. C., Kurdistani, S. K., Morris, S. A., Ahn, S. H., Podolny, V., Collins, S. R., . . . Krogan, N. J. (2005). Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell, 123*(4), 593-605. doi:10.1016/j.cell.2005.10.025

Kestler, H. A., Müller, A., Gress, T. M., & Buchholz, M. (2005). Generalized Venn diagrams: a new method of visualizing complex genetic set relations. *Bioinformatics, 21*(8), 1592-1595. doi:10.1093/bioinformatics/bti169

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol, 37*(8), 907-915. doi:10.1038/s41587-019-0201-4

Kohwi, Y., & Kohwi-Shigematsu, T. (1988). Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proceedings of the National Academy of Sciences of the United States of America, 85*(11), 3781-3785. doi:10.1073/pnas.85.11.3781

Konig, F., Schubert, T., & Langst, G. (2017). The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. *PLoS One, 12*(6), e0178875. doi:10.1371/journal.pone.0178875

Konig, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., . . . Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol, 17*(7), 909-915. doi:10.1038/nsmb.1838

Kopp, F., & Mendell, J. T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell, 172*(3), 393-407. doi:10.1016/j.cell.2018.01.011

Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., . . . Greenblatt, J. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol, 23*(12), 4207-4218. doi:10.1128/mcb.23.12.4207-4218.2003

Kubik, S., Bruzzone, M. J., Jacquet, P., Falcone, J. L., Rougemont, J., & Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. *Mol Cell, 60*(3), 422-434. doi:10.1016/j.molcel.2015.10.002

Kubik, S., Bruzzone, M. J., & Shore, D. (2017). TFIID or not TFIID, a continuing transcriptional SAGA. *EMBO J, 36*(3), 248-249. doi:10.15252/embj.201696152

Kwok, C. K., Ding, Y., Sherlock, M. E., Assmann, S. M., & Bevilacqua, P. C. (2013). A hybridization-based approach for quantitative and low-bias single-stranded DNA ligation. *Anal Biochem, 435*(2), 181-186. doi:10.1016/j.ab.2013.01.008

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., & Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell, 121*(5), 713-724. doi:10.1016/j.cell.2005.04.029

Landgraf, R., Chen, C. H., & Sigman, D. S. (1995a). Double stranded scission of DNA directed through sequence-specific R-loop formation. *Nucleic Acids Res, 23*(17), 3524-3530. doi:10.1093/nar/23.17.3524

Landgraf, R., Chen, C. H., & Sigman, D. S. (1995b). R-loop stability as a function of RNA structure and size. *Nucleic Acids Res, 23*(17), 3516-3523. doi:10.1093/nar/23.17.3516

Landgraf, R., Ramamurthi, K. S., & Sigman, D. S. (1996). Kinetics of spontaneous displacement of RNA from heteroduplexes by DNA. *Nucleic Acids Res, 24*(16), 3246-3252. doi:10.1093/nar/24.16.3246

Landry, J., Sutton, A., Hesman, T., Min, J., Xu, R. M., Johnston, M., & Sternglanz, R. (2003). Set2-Catalyzed Methylation of Histone H3 Represses Basal Expression of GAL4 in Saccharomyces cerevisiae. *Molecular and Cellular Biology, 23*(17), 5972-5978. doi:10.1128/mcb.23.17.5972-5978.2003

Larigot, L., Juricek, L., Dairou, J., & Coumoul, X. (2018). AhR signaling pathways and regulatory functions. *Biochim Open, 7*, 1-9. doi:10.1016/j.biopen.2018.05.001

Lee, D. Y., & Clayton, D. A. (1998). Initiation of mitochondrial DNA replication by transcription and R-loop processing. *J Biol Chem, 273*(46), 30614-30621. doi:10.1074/jbc.273.46.30614

Leela, J. K., Syeda, A. H., Anupama, K., & Gowrishankar, J. (2013). Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. *Proc Natl Acad Sci U S A, 110*(1), 258-263. doi:10.1073/pnas.1213123110

Legros, P., Malapert, A., Niinuma, S., Bernard, P., & 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 Genet, 10*(11), e1004794. doi:10.1371/journal.pgen.1004794

Lesnik, E. A., & Freier, S. M. (1995). Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: relationship with base composition and structure. *Biochemistry, 34*(34), 10807-10815. doi:10.1021/bi00034a013

Li, B., Carey, M., & Workman, J. L. (2007). The role of chromatin during transcription. *Cell, 128*(4), 707-719. doi:10.1016/j.cell.2007.01.015

Li, J., Duns, G., Westers, H., Sijmons, R., van den Berg, A., & Kok, K. (2016). SETD2: an epigenetic modifier with tumor suppressor functionality. *Oncotarget, 7*(31), 50719-50734. doi:10.18632/oncotarget.9368

Li, X., & Manley, J. L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell, 122*(3), 365-378. doi:10.1016/j.cell.2005.06.008

Liang, Z., Liang, F., Teng, Y., Chen, X., Liu, J., Longerich, S., . . . Kupfer, G. M. (2019). Binding of FANCI-FANCD2 Complex to RNA and R-Loops Stimulates Robust FANCD2 Monoubiquitination. *Cell Rep, 26*(3), 564-572 e565. doi:10.1016/j.celrep.2018.12.084

Liu, L. F., & Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci U S A, 84*(20), 7024-7027. doi:10.1073/pnas.84.20.7024

Liu, X., Bushnell, D. A., & Kornberg, R. D. (2013). RNA polymerase II transcription: structure and mechanism. *Biochim Biophys Acta, 1829*(1), 2-8. doi:10.1016/j.bbagrm.2012.09.003

Liu, X., Farnung, L., Wigge, C., & Cramer, P. (2018). Cryo-EM structure of a mammalian RNA polymerase II elongation complex inhibited by alpha-amanitin. *J Biol Chem, 293*(19), 7189-7194. doi:10.1074/jbc.RA118.002545

Lu, B., Dong, L., Yi, D., Zhang, M., Zhu, C., Li, X., & Yi, C. (2020). Transposase-assisted tagmentation of RNA/DNA hybrid duplexes. *Elife, 9*. doi:10.7554/eLife.54919

Madireddy, A., Kosiyatrakul, S. T., Boisvert, R. A., Herrera-Moyano, E., Garcia-Rubio, M. L., Gerhardt, J., . . . Schildkraut, C. L. (2016). FANCD2 Facilitates Replication through Common Fragile Sites. *Mol Cell, 64*(2), 388-404. doi:10.1016/j.molcel.2016.09.017

Maffia, A., Ranise, C., & Sabbioneda, S. (2020). From R-Loops to G-Quadruplexes: Emerging New Threats for the Replication Fork. *Int J Mol Sci, 21*(4). doi:10.3390/ijms21041506

Maizels, N., Groh, M., & Gromak, N. (2014). Out of Balance: R-loops in Human Disease. *PLoS Genetics, 10*(9), e1004630. doi:10.1371/journal.pgen.1004630

Malig, M., Hartono, S. R., Giafaglione, J. M., Sanz, L. A., & Chedin, F. (2020). Ultra-deep Coverage Single-molecule R-loop Footprinting Reveals Principles of R-loop Formation. *J Mol Biol*. doi:10.1016/j.jmb.2020.02.014

Marina, D. B., Shankar, S., Natarajan, P., Finn, K. J., & Madhani, H. D. (2013). A conserved ncRNA-binding protein recruits silencing factors to heterochromatin through an RNAi-independent mechanism. *Genes Dev, 27*(17), 1851-1856. doi:10.1101/gad.226019.113

Martienssen, R., & Moazed, D. (2015). RNAi and heterochromatin assembly. *Cold Spring Harb Perspect Biol, 7*(8), a019323. doi:10.1101/cshperspect.a019323

Massé, E., & Drolet, M. (1999a). Escherichia coli DNA topoisomerase I inhibits R-loop formation by relaxing transcription-induced negative supercoiling. *J Biol Chem, 274*(23), 16659-16664. doi:10.1074/jbc.274.23.16659

Massé, E., & Drolet, M. (1999b). R-loop-dependent hypernegative supercoiling in Escherichia coli topA mutants preferentially occurs at low temperatures and correlates with growth inhibition. *J Mol Biol, 294*(2), 321-332. doi:10.1006/jmbi.1999.3264

McDaniel, S. L., & Strahl, B. D. (2017). Shaping the cellular landscape with Set2/SETD2 methylation. *Cell Mol Life Sci, 74*(18), 3317-3334. doi:10.1007/s00018-017-2517-x

McGhee, J. D., & Von Hippel, P. H. (1977). Formaldehyde as a probe of DNA structure. 4. Mechanism of the initial reaction of formaldehyde with DNA. *Biochemistry, 16*(15), 3276-3293. doi:10.1021/bi00634a002

Meers, M. P., Henriques, T., Lavender, C. A., McKay, D. J., Strahl, B. D., Duronio, R. J., . . . Matera, A. G. (2017). Histone gene replacement reveals a post-transcriptional role for H3K36 in maintaining metazoan transcriptome fidelity. *Elife, 6*. doi:10.7554/eLife.23249

Milman, G., Langridge, R., & Chamberlin, M. J. (1967). The structure of a DNA-RNA hybrid. *Proceedings of the National Academy of Sciences, 57*(6), 1804-1810. doi:10.1073/pnas.57.6.1804

Mirkin, S. M. (2008). Discovery of alternative DNA structures: a heroic decade (1979-1989). *Front Biosci, 13*, 1064-1071. doi:10.2741/2744

Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, V. N., Filippov, S. A., & Frank-Kamenetskii, M. D. (1987). DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature, 330*(6147), 495-497. doi:10.1038/330495a0

Mischo, H. E., Gomez-Gonzalez, B., Grzechnik, P., Rondon, A. G., Wei, W., Steinmetz, L., . . . Proudfoot, N. J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell, 41*(1), 21-32. doi:10.1016/j.molcel.2010.12.007

Mishra, P. K., Chakraborty, A., Yeh, E., Feng, W., Bloom, K. S., & Basrai, M. A. (2021). R-loops at centromeric chromatin contribute to defects in kinetochore integrity and chromosomal instability in budding yeast. *Mol Biol Cell, 32*(1), 74-89. doi:10.1091/mbc.E20-06-0379

Murr, R. (2010). Interplay between different epigenetic modifications and mechanisms. *Adv Genet, 70*, 101-141. doi:10.1016/B978-0-12-380866-0.60005-8

Murray, S. C., Haenni, S., Howe, F. S., Fischl, H., Chocian, K., Nair, A., & Mellor, J. (2015). Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Res, 43*(16), 7823-7837. doi:10.1093/nar/gkv666

Murthy, V. L., Srinivasan, R., Draper, D. E., & Rose, G. D. (1999). A Complete Conformational Map for RNA. *Journal of Molecular Biology, 291*(2), 313-327. doi:https://doi.org/10.1006/jmbi.1999.2958

Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N. A., P, O. B., Sato, H., . . . Greally, J. M. (2015). RNA:DNA hybrids in the human genome have distinctive nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics Chromatin, 8*, 46. doi:10.1186/s13072-015-0040-6

Nakama, M., Kawakami, K., Kajitani, T., Urano, T., & Murakami, Y. (2012). DNA-RNA hybrid formation mediates RNAi-directed heterochromatin formation. *Genes Cells, 17*(3), 218-233. doi:10.1111/j.1365-2443.2012.01583.x

Nakayama, J., Xiao, G., Noma, K., Malikzay, A., Bjerling, P., Ekwall, K., . . . Grewal, S. I. (2003). Alp13, an MRG family protein, is a component of fission yeast Clr6 histone deacetylase required for genomic integrity. *EMBO J, 22*(11), 2776-2787. doi:10.1093/emboj/cdg248

Natsume-Kitatani, Y., & Mamitsuka, H. (2016). Classification of Promoters Based on the Combination of Core Promoter Elements Exhibits Different Histone Modification Patterns. *PLoS One, 11*(3), e0151917. doi:10.1371/journal.pone.0151917

Neil, A. J., Liang, M. U., Khristich, A. N., Shah, K. A., & Mirkin, S. M. (2018). RNA-DNA hybrids promote the expansion of Friedreich's ataxia (GAA)n repeats via break-induced replication. *Nucleic Acids Res, 46*(7), 3487-3497. doi:10.1093/nar/gky099

Nicolas, E., Yamada, T., Cam, H. P., Fitzgerald, P. C., Kobayashi, R., & Grewal, S. I. (2007). Distinct roles of HDAC complexes in promoter silencing, antisense suppression and DNA damage protection. *Nat Struct Mol Biol, 14*(5), 372-380. doi:10.1038/nsmb1239

Niehrs, C., & Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol, 21*(3), 167-178. doi:10.1038/s41580-019-0206-3

Noy, A., Pérez, A., Márquez, M., Luque, F. J., & Orozco, M. (2005). Structure, Recognition Properties, and Flexibility of the DNA·RNA Hybrid. *Journal of the American Chemical Society, 127*(13), 4910-4920. doi:10.1021/ja043293v

Nudler, E. (2012). RNA polymerase backtracking in gene regulation and genome instability. *Cell, 149*(7), 1438-1445. doi:10.1016/j.cell.2012.06.003

O'Brien, E. J., & MacEwan, A. W. (1970). Molecular and crystal structure of the polynucleotide complex: polyinosinic acid plus polydeoxycytidylic acid. *Journal of Molecular Biology, 48*(2), 243-261. doi:10.1016/0022-2836(70)90159-2

Ohle, C., Tesorero, R., Schermann, G., Dobrev, N., Sinning, I., & Fischer, T. (2016). Transient RNA-DNA Hybrids Are Required for Efficient Double-Strand Break Repair. *Cell, 167*(4), 1001-1013 e1007. doi:10.1016/j.cell.2016.10.001

Okamoto, Y., Abe, M., Itaya, A., Tomida, J., Ishiai, M., Takaori-Kondo, A., . . . Takata, M. (2019). FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. *FEBS J, 286*(1), 139-150. doi:10.1111/febs.14700

Pan, G., & Greenblatt, J. (1994). Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site. *Journal of Biological Chemistry, 269*(48), 30101-30104. doi:10.1016/s0021-9258(18)43780-5

Pardi, A., Martin, F. H., & Tinoco, I. (1981). Comparative study of ribonucleotides, deoxyribonucleotides, and hybrid oligonucleotide helixes by nuclear magnetic resonance. *Biochemistry, 20*(14), 3986-3996. doi:10.1021/bi00517a007

Parsa, J. Y., Boudoukha, S., Burke, J., Homer, C., & Madhani, H. D. (2018). Polymerase pausing induced by sequence-specific RNA-binding protein drives heterochromatin assembly. *Genes Dev, 32*(13-14), 953-964. doi:10.1101/gad.310136.117

Parvin, J. D., & Sharp, P. A. (1993). DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell, 73*(3), 533-540. doi:10.1016/0092-8674(93)90140-l

Pattenden, S. G., Gogol, M. M., & Workman, J. L. (2010). Features of cryptic promoters and their varied reliance on bromodomain-containing factors. *PLoS One, 5*(9), e12927. doi:10.1371/journal.pone.0012927

Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., . . . Basu, U. (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell, 161*(4), 774-789. doi:10.1016/j.cell.2015.04.034

Peng, X., Wu, J., Brunmeir, R., Kim, S. Y., Zhang, Q., Ding, C., . . . Xu, F. (2015). TELP, a sensitive and versatile library construction method for next-generation sequencing. *Nucleic Acids Res, 43*(6), e35. doi:10.1093/nar/gku818

Peng, Z., Yuan, C., Zellmer, L., Liu, S., Xu, N., & Liao, D. J. (2015). Hypothesis: Artifacts, Including Spurious Chimeric RNAs with a Short Homologous Sequence, Caused by Consecutive Reverse Transcriptions and Endogenous Random Primers. *J Cancer, 6*(6), 555-567. doi:10.7150/jca.11997

Perez-Martinez, L., Ozturk, M., Butter, F., & Luke, B. (2020). Npl3 stabilizes R-loops at telomeres to prevent accelerated replicative senescence. *EMBO Rep, 21*(3), e49087. doi:10.15252/embr.201949087

Peters, J. M., Mooney, R. A., Grass, J. A., Jessen, E. D., Tran, F., & Landick, R. (2012). Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev, 26*(23), 2621-2633. doi:10.1101/gad.196741.112

Peters, J. M., Vangeloff, A. D., & Landick, R. (2011). Bacterial transcription terminators: the RNA 3'-end chronicles. *J Mol Biol, 412*(5), 793-813. doi:10.1016/j.jmb.2011.03.036

Phillips, D. D., Garboczi, D. N., Singh, K., Hu, Z., Leppla, S. H., & 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*(8), 376-381. doi:10.1002/jmr.2284

Picelli, S., Bjorklund, A. K., Faridani, O. R., Sagasser, S., Winberg, G., & Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods, 10*(11), 1096-1098. doi:10.1038/nmeth.2639

Picelli, S., Faridani, O. R., Bjorklund, A. K., Winberg, G., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc, 9*(1), 171-181. doi:10.1038/nprot.2014.006

Pointner, J., Persson, J., Prasad, P., Norman-Axelsson, U., Stralfors, A., Khorosjutina, O., . . . Korber, P. (2012). CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. *EMBO J, 31*(23), 4388-4403. doi:10.1038/emboj.2012.289

Porrua, O., & Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol, 20*(7), 884-891. doi:10.1038/nsmb.2592

Powell, W. T., Coulson, R. L., Gonzales, M. L., Crary, F. K., Wong, S. S., Adams, S., . . . LaSalle, J. M. (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proc Natl Acad Sci U S A, 110*(34), 13938-13943. doi:10.1073/pnas.1305426110

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. doi:10.1093/bioinformatics/btq033

Rabhi, M., Rahmouni, A. R., & Boudvillain, M. (2010). Chapter 10. Transcription Termination Factor Rho: A Ring-Shaped RNA Helicase from Bacteria. In *RNA Helicases* (pp. 243-271).

Raghunathan, N., Kapshikar, R. M., Leela, J. K., Mallikarjun, J., Bouloc, P., & Gowrishankar, J. (2018). Genome-wide relationship between R-loop formation and antisense transcription in Escherichia coli. *Nucleic Acids Res, 46*(7), 3400-3411. doi:10.1093/nar/gky118

Ragunathan, K., Jih, G., & Moazed, D. (2015). Epigenetics. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science, 348*(6230), 1258699. doi:10.1126/science.1258699

Raine, A., Manlig, E., Wahlberg, P., Syvanen, A. C., & Nordlund, J. (2017). SPlinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. *Nucleic Acids Res, 45*(6), e36. doi:10.1093/nar/gkw1110

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., . . . Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res, 44*(W1), W160-165. doi:10.1093/nar/gkw257

Rando, O. J., & Chang, H. Y. (2009). Genome-wide views of chromatin structure. *Annu Rev Biochem, 78*, 245-271. doi:10.1146/annurev.biochem.78.071107.134639

Ransohoff, J. D., Wei, Y., & Khavari, P. A. (2018). The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol, 19*(3), 143-157. doi:10.1038/nrm.2017.104

Ratmeyer, L., Vinayak, R., Zhong, Y. Y., Zon, G., & Wilson, W. D. (1994). Sequence specific thermodynamic and structural properties for DNA.RNA duplexes. *Biochemistry, 33*(17), 5298-5304. doi:10.1021/bi00183a037

Reaban, M. E., Lebowitz, J., & Griffin, J. A. (1994). Transcription induces the formation of a stable RNA.DNA hybrid in the immunoglobulin alpha switch region. *J Biol Chem, 269*(34), 21850-21857.

Recillas-Targa, F., & Razin, S. V. (2001). Chromatin domains and regulation of gene expression: familiar and enigmatic clusters of chicken globin genes. *Crit Rev Eukaryot Gene Expr, 11*(1-3), 227-242.

Reyes-Turcu, F. E., Zhang, K., Zofall, M., Chen, E., & Grewal, S. I. (2011). Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. *Nat Struct Mol Biol, 18*(10), 1132-1138. doi:10.1038/nsmb.2122

Rhee, H. S., & Pugh, B. F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell, 147*(6), 1408-1419. doi:10.1016/j.cell.2011.11.013

Rhee, H. S., & Pugh, B. F. (2012). ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol, Chapter 21*, Unit 21 24. doi:10.1002/0471142727.mb2124s100

Rivosecchi, J., Larochelle, M., Teste, C., Grenier, F., Malapert, A., Ricci, E. P., . . . Vanoosthuyse, V. (2019). Senataxin homologue Sen1 is required for efficient termination of RNA polymerase III transcription. *EMBO J, 38*(16), e101955. doi:10.15252/embj.2019101955

Robberson, D. L., Kasamatsu, H., & Vinograd, J. (1972). Replication of Mitochondrial DNA. Circular Replicative Intermediates

in Mouse L Cells. *Proc. Nat. Acad. Sci., 69*(3 ), 737-741.

Roberts, R. W., & Crothers, D. M. (1992). Stability and Properties of Double and Triple Helices: Dramatic Effects of RNA or DNA Backbone Composition. *Science 258*(5087), 1463-1466. doi:DOI: 10.1126/science.1279808

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nat Biotechnol, 29*(1), 24-26. doi:10.1038/nbt.1754

Rodríguez-Navarro, S., Fischer, T., Luo, M. J., Antúnez, O., Brettschneider, S., Lechner, J., . . . Hurt, E. (2004). Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell, 116*(1), 75-86. doi:10.1016/s0092-8674(03)01025-0

Rossi, M. J., Lai, W. K. M., & Pugh, B. F. (2018). Simplified ChIP-exo assays. *Nat Commun, 9*(1), 2842. doi:10.1038/s41467-018-05265-7

Rosso, I., & d'Adda di Fagagna, F. (2020). Detection of Telomeric DNA:RNA Hybrids Using TeloDRIP-qPCR. *Int J Mol Sci, 21*(24). doi:10.3390/ijms21249774

Roy, D., & Lieber, M. R. (2009). G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol, 29*(11), 3124-3133. doi:10.1128/MCB.00139-09

Roy, D., Yu, K., & Lieber, M. R. (2008). Mechanism of R-loop formation at immunoglobulin class switch sequences. *Mol Cell Biol, 28*(1), 50-60. doi:10.1128/MCB.01251-07

Roy, D., Zhang, Z., Lu, Z., Hsieh, C. L., & 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. *Mol Cell Biol, 30*(1), 146-159. doi:10.1128/mcb.00897-09

Rudolph, C. J., Upton, A. L., Briggs, G. S., & Lloyd, R. G. (2010). Is RecG a general guardian of the bacterial genome? *DNA Repair (Amst), 9*(3), 210-223. doi:10.1016/j.dnarep.2009.12.014

Sadhu, C., Dutta, S., & Gopinathan, K. P. (1984). Influence of formamide on the thermal stability of DNA. *Journal of Biosciences, 6*(6), 817-821. doi:10.1007/BF02716841

Saenger, W. (1984). Polymorphism of DNA versus Structural Conservatism of RNA: Classification of A-, B-, and Z-TYPe Double Helices. In *Principles of Nucleic Acid Structure.* (pp. 220–241): Springer, New York, NY.

Salazar, M., Fedoroff, O. Y., Miller, J. M., Ribeiro, N. S., & Reid, B. R. (1993). The DNA strand in DNA.RNA hybrid duplexes is neither B-form nor A-form in solution. *Biochemistry, 32*(16), 4207-4215. doi:10.1021/bi00067a007

Santos-Pereira, J. M., & Aguilera, A. (2015). R loops: new modulators of genome dynamics and function. *Nat Rev Genet, 16*(10), 583-597. doi:10.1038/nrg3961

Sanz, L. A., & Chedin, F. (2019). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. *Nat Protoc, 14*(6), 1734-1755. doi:10.1038/s41596-019-0159-1

Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., . . . Chedin, F. (2016). Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific

Epigenomic Signatures in Mammals. *Mol Cell, 63*(1), 167-178. doi:10.1016/j.molcel.2016.05.032

Schier, A. C., & Taatjes, D. J. (2020). Structure and mechanism of the RNA polymerase II transcription machinery. *Genes Dev, 34*, 465–488. doi:10.1101/gad.335679

Schmitz, K. M., Mayer, C., Postepska, A., & Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev, 24*(20), 2264-2269. doi:10.1101/gad.590910

Schneider, B., Moravek, Z., & Berman, H. M. (2004). RNA conformational classes. *Nucleic Acids Res, 32*(5), 1666-1677. doi:10.1093/nar/gkh333

Schwab, R. A., Nieminuszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C. C., . . . Niedzwiedz, W. (2015). The Fanconi Anemia Pathway Maintains Genome Stability by Coordinating Replication and Transcription. *Mol Cell, 60*(3), 351-361. doi:10.1016/j.molcel.2015.09.012

Sharma, S., Langhendries, J. L., Watzinger, P., Kötter, P., Entian, K. D., & Lafontaine, D. L. (2015). Yeast Kre33 and human NAT10 are conserved 18S rRNA cytosine acetyltransferases that modify tRNAs assisted by the adaptor Tan1/THUMPD1. *Nucleic Acids Res, 43*(4), 2242-2258. doi:10.1093/nar/gkv075

Shaw, N. N., & Arya, D. P. (2008). Recognition of the unique structure of DNA:RNA hybrids. *Biochimie, 90*(7), 1026-1039. doi:10.1016/j.biochi.2008.04.011

Shen, L., Shao, N., Liu, X., & Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics, 15*, 284. doi:10.1186/1471-2164-15-284

Shichino, Y., Otsubo, Y., Yamamoto, M., & Yamashita, A. (2020). Meiotic gene silencing complex MTREC/NURS recruits the nuclear exosome to YTH-RNA-binding protein Mmi1. *PLoS Genet, 16*(2), e1008598. doi:10.1371/journal.pgen.1008598

Shim, Y. S., Choi, Y., Kang, K., Cho, K., Oh, S., Lee, J., . . . Lee, D. (2012). Hrp3 controls nucleosome positioning to suppress non-coding transcription in eu- and heterochromatin. *EMBO J, 31*(23), 4375-4387. doi:10.1038/emboj.2012.267

Shindo, H., & Matsumoto, U. (1984). Direct evidence for a bimorphic structure of a DNA-RNA hybrid, poly(rA).poly(dT), at high relative humidity. *J Biol Chem, 259*(14), 8682-8684.

Skene, P. J., Henikoff, J. G., & Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc, 13*(5), 1006-1019. doi:10.1038/nprot.2018.015

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife, 6*. doi:10.7554/eLife.21856

Skourti-Stathaki, K., Kamieniarz-Gdula, K., & Proudfoot, N. J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature, 516*(7531), 436-439. doi:10.1038/nature13787

Skourti-Stathaki, K., & Proudfoot, N. J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev, 28*(13), 1384-1396. doi:10.1101/gad.242990.114

Skourti-Stathaki, K., Proudfoot, N. J., & Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell, 42*(6), 794-805. doi:10.1016/j.molcel.2011.04.026

Skourti-Stathaki, K., Torlai Triglia, E., Warburton, M., Voigt, P., Bird, A., & Pombo, A. (2019). R-Loops Enhance Polycomb Repression at a Subset of Developmental Regulator Genes. *Mol Cell*. doi:10.1016/j.molcel.2018.12.016

Smolka, J. A., Sanz, L. A., Hartono, S. R., & Chédin, F. (2021). Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. *J Cell Biol, 220*(6). doi:10.1083/jcb.202004079

Smolle, M., Venkatesh, S., Gogol, M. M., Li, H., Zhang, Y., Florens, L., . . . Workman, J. L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat Struct Mol Biol, 19*(9), 884-892. doi:10.1038/nsmb.2312

Sorenson, M. R., Jha, D. K., Ucles, S. A., Flood, D. M., Strahl, B. D., Stevens, S. W., & Kress, T. L. (2016). Histone H3K36 methylation regulates pre-mRNA splicing in Saccharomyces cerevisiae. *RNA Biol, 13*(4), 412-426. doi:10.1080/15476286.2016.1144009

Statello, L., Guo, C. J., Chen, L. L., & Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol, 22*(2), 96-118. doi:10.1038/s41580-020-00315-9

Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z. W., Bone, J. R., Caldwell, J. A., . . . Allis, C. D. (2002). Set2 Is a Nucleosomal Histone H3-Selective Methyltransferase That Mediates Transcriptional Repression. *Molecular and Cellular Biology, 22*(5), 1298-1306. doi:10.1128/mcb.22.5.1298-1306.2002

Strässer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., . . . Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. *Nature, 417*(6886), 304-308. doi:10.1038/nature746

Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., . . . Sasaki, M. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry, 34*(35), 11211-11216. doi:10.1021/bi00035a029

Sugiyama, T., Thillainadesan, G., Chalamcharla, V. R., Meng, Z., Balachandran, V., Dhakshnamoorthy, J., . . . Grewal, S. I. S. (2016). Enhancer of Rudimentary Cooperates with Conserved RNA-Processing Factors to Promote Meiotic mRNA Decay and Facultative Heterochromatin Assembly. *Mol Cell, 61*(5), 747-759. doi:10.1016/j.molcel.2016.01.029

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J., & Dean, C. (2013). R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *SCIENCE, 340*(6132), 619-621. doi:10.1126/science.1234848

Suzuki, S., Kato, H., Suzuki, Y., Chikashige, Y., Hiraoka, Y., Kimura, H., . . . Murakami, Y. (2016). Histone H3K36 trimethylation is essential for multiple silencing mechanisms in fission yeast. *Nucleic Acids Res, 44*(9), 4147-4162. doi:10.1093/nar/gkw008

Svozil, D., Kalina, J., Omelka, M., & Schneider, B. (2008). DNA conformations and their sequence preferences. *Nucleic Acids Res, 36*(11), 3690-3706. doi:10.1093/nar/gkn260

Szabat, M., & Kierzek, R. (2017). Parallel-stranded DNA and RNA duplexes - structural features and potential applications. *FEBS J, 284*(23), 3986-3998. doi:10.1111/febs.14187

Tafur, L., Sadian, Y., Hoffmann, N. A., Jakobi, A. J., Wetzel, R., Hagen, W. J. H., . . . Muller, C. W. (2016). Molecular Structures of Transcribing RNA Polymerase I. *Mol Cell, 64*(6), 1135-1143. doi:10.1016/j.molcel.2016.11.013

Tan-Wong, S. M., Dhir, S., & Proudfoot, N. J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol Cell, 76*(4), 600-616 e606. doi:10.1016/j.molcel.2019.10.002

Tateishi-Karimata, H., & Sugimoto, N. (2014). Structure, stability and behaviour of nucleic acids in ionic liquids. *Nucleic Acids Res, 42*(14), 8831-8844. doi:10.1093/nar/gku499

Thillainadesan, G., Xiao, H., Holla, S., Dhakshnamoorthy, J., Jenkins, L. M. M., Wheeler, D., & Grewal, S. I. S. (2020). Conserved protein Pir2(ARS2) mediates gene repression through cryptic introns in lncRNAs. *Nat Commun, 11*(1), 2412. doi:10.1038/s41467-020-16280-y

Thomas, M., White, R. L., & Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: Formation of

R-loops. *Proc. Natl. Acad. Sci., 73*(7), 2294-2298.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform, 14*(2), 178-192. doi:10.1093/bib/bbs017

Till, P., Mach, R. L., & Mach-Aigner, A. R. (2018). A current view on long noncoding RNAs in yeast and filamentous fungi. *Appl Microbiol Biotechnol, 102*(17), 7319-7331. doi:10.1007/s00253-018-9187-y

Touat-Todeschini, L., Hiriart, E., & Verdel, A. (2012). Nucleosome positioning and transcription: fission yeast CHD remodellers make their move. *EMBO J, 31*(23), 4371-4372. doi:10.1038/emboj.2012.284

Touat-Todeschini, L., Shichino, Y., Dangin, M., Thierry-Mieg, N., Gilquin, B., Hiriart, E., . . . Verdel, A. (2017). Selective termination of lncRNA transcription promotes heterochromatin silencing and cell differentiation. *EMBO J, 36*(17), 2626-2641. doi:10.15252/embj.201796571

Tous, C., & Aguilera, A. (2007). Impairment of transcription elongation by R-loops in vitro. *Biochem Biophys Res Commun, 360*(2), 428-432. doi:10.1016/j.bbrc.2007.06.098

Turchinovich, A., Surowy, H., Serva, A., Zapatka, M., Lichter, P., & Burwinkel, B. (2014). Capture and Amplification by Tailing and Switching (CATS). An ultrasensitive ligation-independent method for generation of DNA libraries for deep sequencing from picogram amounts of DNA and RNA. *RNA Biol, 11*(7), 817-828. doi:10.4161/rna.29304

Vanácová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., . . . Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol, 3*(6), e189. doi:10.1371/journal.pbio.0030189

Vanoosthuyse, V. (2018). Strengths and Weaknesses of the Current Strategies to Map and Characterize R-Loops. *Noncoding RNA, 4*(2). doi:10.3390/ncrna4020009

Vardi, O., Shamir, I., Javasky, E., Goren, A., & Simon, I. (2017). Biases in the SMART-DNA library preparation method associated with genomic poly dA/dT sequences. *PLoS One, 12*(2), e0172769. doi:10.1371/journal.pone.0172769

Venkatesh, S., Li, H., Gogol, M. M., & Workman, J. L. (2016). Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nat Commun, 7*, 13610. doi:10.1038/ncomms13610

Venkatesh, S., Smolle, M., Li, H., Gogol, M. M., Saint, M., Kumar, S., . . . Workman, J. L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature, 489*(7416), 452-455. doi:10.1038/nature11326

Venkatesh, S., & Workman, J. L. (2013). Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. *Wiley Interdiscip Rev Dev Biol, 2*(5), 685-700. doi:10.1002/wdev.109

Venkatesh, S., & Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol, 16*(3), 178-189. doi:10.1038/nrm3941

Verboom, K., Everaert, C., Bolduc, N., Livak, K. J., Yigit, N., Rombaut, D., . . . Vandesompele, J. (2019). SMARTer single cell total RNA sequencing. *Nucleic Acids Res, 47*(16), e93. doi:10.1093/nar/gkz535

Vo, T. V., Dhakshnamoorthy, J., Larkin, M., Zofall, M., Thillainadesan, G., Balachandran, V., . . . Grewal, S. I. S. (2019). CPF Recruitment to Non-canonical Transcription Termination Sites Triggers Heterochromatin Assembly and Gene Silencing. *Cell Rep, 28*(1), 267-281 e265. doi:10.1016/j.celrep.2019.05.107

Volkmann, S., Jendis, J., Frauendorf, A., & Moelling, K. (1995). Inhibition of HIV-1 reverse transcription by triple-helix forming oligonucleotides with viral RNA. *Nucleic Acids Res, 23*(7), 1204-1212. doi:10.1093/nar/23.7.1204

Wagner, E. J., & Carpenter, P. B. (2012). Understanding the language of Lys36 methylation at histone H3. *Nat Rev Mol Cell Biol, 13*(2), 115-126. doi:10.1038/nrm3274

Wahba, L., Amon, J. D., Koshland, D., & Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell, 44*(6), 978-988. doi:10.1016/j.molcel.2011.10.017

Wahba, L., Costantino, L., Tan, F. J., Zimmer, A., & Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev, 30*(11), 1327-1338. doi:10.1101/gad.280834.116

Wahba, L., Gore, S. K., & Koshland, D. (2013). The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *Elife, 2*, e00505. doi:10.7554/eLife.00505

Wahba, L., & Koshland, D. (2013). The Rs of biology: R-loops and the regulation of regulators. *Mol Cell, 50*(5), 611-612. doi:10.1016/j.molcel.2013.05.024

Wang, A. H., Fujii, S., van Boom, J. H., van der Marel, G. A., van Boeckel, S. A., & Rich, A. (1982). Molecular structure of r(GCG)d(TATACGC): a DNA--RNA hybrid helix joined to double helical DNA. *Nature, 299*(5884), 601-604. doi:10.1038/299601a0

Wang, G., & Vasquez, K. M. (2014). Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability. *DNA Repair (Amst), 19*, 143-151. doi:10.1016/j.dnarep.2014.03.017

Wang, G., & Vasquez, K. M. (2017). Effects of Replication and Transcription on DNA Structure-Related Genetic Instability. *Genes (Basel), 8*(1). doi:10.3390/genes8010017

Wang, J. C. (1974). Interactions between twisted DNAs and enzymes: the effects of superhelical turns. *J Mol Biol, 87*(4), 797-816. doi:10.1016/0022-2836(74)90085-0

Wang, K., Wang, H., Li, C., Yin, Z., Xiao, R., Li, Q., . . . Liang, K. (2021). Genomic profiling of native R loops with a DNA-RNA hybrid recognition sensor. *Sci Adv, 7*(8). doi:10.1126/sciadv.abe3516

Wang, X., & Moazed, D. (2017). DNA sequence-dependent epigenetic inheritance of gene silencing and histone H3K9 methylation. *Science, 356*(6333), 88-91. doi:10.1126/science.aaj2114

Watanabe, K., & Kokubo, T. (2017). SAGA mediates transcription from the TATA-like element independently of Taf1p/TFIID but dependent on core promoter structures in Saccharomyces cerevisiae. *PLoS One, 12*(11), e0188435. doi:10.1371/journal.pone.0188435

Wei, W., Hennig, B. P., Wang, J., Zhang, Y., Piazza, I., Pareja Sanchez, Y., . . . Pelechano, V. (2019). Chromatin-sensitive cryptic promoters putatively drive expression of alternative protein isoforms in yeast. *Genome Res, 29*(12), 1974-1984. doi:10.1101/gr.243378.118

Wery, M., Descrimes, M., Vogt, N., Dallongeville, A. S., Gautheret, D., & Morillon, A. (2016). Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Mol Cell, 61*(3), 379-392. doi:10.1016/j.molcel.2015.12.020

Wery, M., Gautier, C., Descrimes, M., Yoda, M., Migeot, V., Hermand, D., & Morillon, A. (2018). Bases of antisense lncRNA-associated regulation of gene expression in fission yeast. *PLoS Genet, 14*(7), e1007465. doi:10.1371/journal.pgen.1007465

Wery, M., Gautier, C., Descrimes, M., Yoda, M., Vennin-Rendos, H., Migeot, V., . . . Morillon, A. (2018). Native elongating transcript sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast. *RNA (New York, N.Y.), 24*(2), 196-208. doi:10.1261/rna.063446.117

Westover, K. D., Bushnell, D. A., & Kornberg, R. D. (2004). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *SCIENCE, 303*(5660), 1014-1016. doi:10.1126/science.1090839

White, R. L., & Hogness, D. S. (1977). R loop mapping of the 18S and 28S sequences in the long and short repeating units of drosophila melanogaster rDNA. *Cell, 10*, 177-192.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.*: Springer-Verlag New York.

Wu, H.-Y., Shyy, S., Wang, J. C., & Liu, L. F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell, 53*(3), 433-440. doi:https://doi.org/10.1016/0092-8674(88)90163-8

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., . . . Jacquier, A. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway

involving a new poly(A) polymerase. *Cell, 121*(5), 725-737. doi:10.1016/j.cell.2005.04.030

Xiao, T., Hall, H., Kizer, K. O., Shibata, Y., Hall, M. C., Borchers, C. H., & Strahl, B. D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes Dev, 17*(5), 654-663. doi:10.1101/gad.1055503

Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., & Sun, Q. (2017). The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat Plants, 3*(9), 704-714. doi:10.1038/s41477-017-0004-x

Yan, Q., Shields, E. J., Bonasio, R., & Sarma, K. (2019). Mapping Native R-Loops Genome-wide Using a Targeted Nuclease Approach. *Cell Rep, 29*(5), 1369-1380 e1365. doi:10.1016/j.celrep.2019.09.052

Ylmaz, M., Ozic, C., & Gok, l. (2012). Principles of Nucleic Acid Separation by Agarose Gel Electrophoresis. In *Gel Electrophoresis - Principles and Basics*.

Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., & Lieber, M. R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol, 4*(5), 442-451. doi:10.1038/ni919

Yu, R., Wang, X., & Moazed, D. (2018). Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. *Nature, 558*(7711), 615-619. doi:10.1038/s41586-018-0239-3

Zaitsev, E. N., & Kowalczykowski, S. C. (2000). A novel pairing process promoted by *Escherichia coli* RecA protein: inverse DNA and RNA strand exchange. *Genes Dev, 14*, 740–749.

Zhang, K., Fischer, T., Porter, R. L., Dhakshnamoorthy, J., Zofall, M., Zhou, M., . . . Grewal, S. I. (2011). Clr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. *SCIENCE, 331*(6024), 1624-1627. doi:10.1126/science.1198712

Zhang, K., Mosch, K., Fischle, W., & Grewal, S. I. (2008). Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat Struct Mol Biol, 15*(4), 381-388. doi:10.1038/nsmb.1406

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol, 9*(9), R137. doi:10.1186/gb-2008-9-9-r137

Zhang, Z. Z., Pannunzio, N. R., Hsieh, C. L., Yu, K., & Lieber, M. R. (2015). Complexities due to single-stranded RNA during antibody detection of genomic rna:dna hybrids. *BMC Res Notes, 8*, 127. doi:10.1186/s13104-015-1092-1

Zhao, D., & Zheng, D. (2018). SMARTcleaner: identify and clean off-target signals in SMART ChIP-seq analysis. *BMC Bioinformatics, 19*(1), 544. doi:10.1186/s12859-018-2577-4

Zhou, Y., Zhu, J., Schermann, G., Ohle, C., Bendrin, K., Sugioka-Sugiyama, R., . . . Fischer, T. (2015). The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun, 6*, 7050. doi:10.1038/ncomms8050

Zhu, K., Lei, P. J., Ju, L. G., Wang, X., Huang, K., Yang, B., . . . Wu, M. (2017). SPOP-containing complex regulates SETD2 stability and H3K36me3-coupled alternative splicing. *Nucleic Acids Res, 45*(1), 92-105. doi:10.1093/nar/gkw814

Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R., & Siebert, P. D. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques, 30*(4), 892-897. doi:10.2144/01304pf02

Zhumabayeva, B., Diatchenko, L., Chenchik, A., & Siebert, P. D. (2001). Use of SMART-generated cDNA for gene expression studies in multiple human tumors. *Biotechniques, 30*(1), 158-163. doi:10.2144/01301pf01

Zilio, N., Wehrkamp-Richter, S., & Boddy, M. N. (2012). A new versatile system for rapid control of gene expression in the fission yeast Schizosaccharomyces pombe. *Yeast, 29*(10), 425-434. doi:10.1002/yea.2920

Zimmerman, S. B., & Pheiffer, B. H. (1981). A RNA.DNA hybrid that can adopt two conformations: an x-ray diffraction study of poly(rA).poly(dT) in concentrated solution or in fibers. *Proc Natl Acad Sci U S A, 78*(1), 78-82. doi:10.1073/pnas.78.1.78

Zofall, M., Yamanaka, S., Reyes-Turcu, F. E., Zhang, K., Rubin, C., & Grewal, S. I. (2012). RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. *Science, 335*(6064), 96-100. doi:10.1126/science.1211651