

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Mechanisms and Interplay of RNA Processing Pathways

**Permalink**

<https://escholarship.org/uc/item/26b7r89c>

**Author**

Barr, Keaton

**Publication Date**

2025

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Mechanisms and Interplay of RNA Processing Pathways

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of  
Philosophy in Biochemistry, Molecular, and Structural Biology

by

Keaton Barr

2025

© Copyright by

Keaton Barr

2025

# ABSTRACT OF THE DISSERTATION

## Mechanisms and Interplay of RNA Processing Pathways

by

Keaton Barr

Doctor of Philosophy in Biochemistry, Molecular, and Structural Biology

University of California, Los Angeles, 2025

Professor Guillaume Chanfreau, Chair

Production of RNAs, their processing, and the regulation thereof encompasses numerous biological pathways that work both on their own and in concert with one another to ensure proper gene expression and cellular function. This thesis will present four works, each exploring various RNA processing pathways, how they function, and how they may interconnect with one another in the production or processing of their given substrate RNAs.

The first work presented examines the relationship between splicing and 3' end processing in *S. cerevisiae*, processes that have been shown to be coupled in mammalian cells. To further explore this relationship, we set out to study the production of RNA in conditions where the cleavage and polyadenylation (CPA) or splicing pathways were inactivated through nuclear depletion of related proteins. We find that unlike in mammalian cells, the processes of splicing and 3' end processing are not directly coupled in *S. cerevisiae*. However, we demonstrate that

in the event of transcriptional read-through, many yeast mRNAs present extended 5' ends due to termination failure of an upstream gene. These 5' extensions are correlated with decreased splicing efficiency in comparison to unextended mRNAs in a length dependent manner.

Furthermore, we find that in the event of transcriptional read-through due to CPA inactivation, novel intergenic and intragenic splicing events may occur utilizing splice sites found within the extended RNA products.

Continuing this work exploring the effects of CPA inactivation on RNA production, we determined the existence of a population of cleavage and polyadenylation independent mRNAs, which are stably produced in the absence CPA. We provide evidence that this population of mRNAs may utilize novel mechanisms for 3' end formation without the primary CPA components. First, we demonstrate that in the absence of CPA, transcriptional roadblocks are capable of producing stable mRNA 3' ends. We also provide evidence for a novel mechanism of 3' end formation which relies on genetically encoded poly(A) tracts. These poly(A) tracts are actively transcribed and correspond to the exact 3' end of their given mRNAs, suggesting they may act as pseudo poly(A) tails.

In the third work, we follow up on previous data from our lab, which established the use of RNase-III mediated decay (RMD) as a pathway utilized for the regulation of bromo-domain factor 2 (*BDF2*) mRNAs. The exact mechanism through which this RMD regulation occurred, however was not yet understood. We show that this RMD dependent degradation of *BDF2* mRNAs is hyperactivated in conditions which are known to induce increased nuclear retention of global mRNAs, such as salt stress or nuclear depletion of 3' end processing factors.

Strikingly, this RMD dependent processing of *BDF2* mRNAs can be prevented in these same conditions through the nuclear depletion of yeast RNase-III, Rnt1p. Together, these data suggest that the regulation of *BDF2* by RMD is result of increased proximity with Rnt1p in the nucleus as a result of increased nuclear retention in stress conditions.

Finally, we explore the relationships between proper splicing and the production of intron-encoded snoRNAs. We present evidence that when splicing is inactivated, a novel hybrid mRNA-snoRNA (hmsnoRNA) species is generated. Here, we explore the mechanisms which regulate the production and decay of these hmsnoRNAs, shedding light on the importance of proper splicing in the production of intron-encoded snoRNAs.

Altogether, the research presented within this thesis furthers the understanding of the relationships between various RNA processing pathways and how these pathways may or may not come together for their proper function in the production of RNAs in the cell. Additionally, we elucidate the existence of potential novel RNA processing pathways in *S. cerevisiae*.

The dissertation of Keaton Barr is approved.

Steven G. Clarke

Albert J. Courey

Tracy L. Johnson

Guillaume Chanfreau, Committee Chair

University of California, Los Angeles

2025

## DEDICATIONS

This thesis is dedicated to my parents, without whom I never would have made it to this point in my life; to my partner, Kellyn for being my support system over these years; and to Scuba, the fat cat who never forgets to remind us when dinner time is.

## THESIS FORMAT AND TABLE OF CONTENTS

The main contents of this thesis will be broken into several parts. Unpublished work will be described in detail here. However, some portions of this thesis work have already been published. In the corresponding chapters, these previously published works will be reprinted as is permitted by Creative Commons Attribution 4.0 International (CC BY 4.0).

**Part 1: Introduction-** An introduction to the main biological processes being studied and laying the groundwork for research performed.....**Pages 1-13**

**Part 2: Transcription termination is critical for splicing efficiency and fidelity-** A preprint manuscript (at the time of this thesis) exploring the connection between 3'-end processing and splicing in *S. cerevisiae*. My contributions to this work constitute all of the cell culturing, sample preparation, NanoPore sequencing and a majority of the RT-PCR. This work could not have been accomplished without the efforts of Kevin He, who performed all the computational work on the resulting NanoPore data I produced; and Andreas J Krumbain, who assisted in performing the dozens of RT-PCR experiments required for this work.....**Pages: 14-40**

**Part 3: Novel pathways of mRNA 3'-end formation: independence from CPA machinery-** Unpublished work exploring the existence and potential mechanisms of a CPA independent mRNA 3'-end processing pathway in *S. cerevisiae*. My contributions to this work constitute all of the cell culturing, sample preparation, NanoPore sequencing, as well as all of the traditional wet-lab experimentation such as Northern Blotting and 3'-RACE. Again, instrumental to this work are the efforts of Kevin He, who handled all of the computational efforts on the NanoPore sequencing data I produced.....**Pages: 41-58**

**Part 4: Stress-induced inhibition of mRNA export triggers RNase III-mediated decay of the *BDF2* mRNA-** Published work exploring the mechanism through the RMD degradation pathway for *BDF2* occurs. A large portion of this work was spearheaded by Charles Wang, who

produced the initial draft and first attempts at most experimental work. My contributions to this work was the replication of the majority of the experiments and most of the follow-up experimentation requested by reviewers prior to publication with the assistance of Dean Neutel. I also performed all of the RNA FISH microscopy in this work.....**Pages 59-71**

**Part 5: Splicing inactivation generates hybrid mRNA-snoRNA transcripts targeted by cytoplasmic RNA decay-** Published work exploring the effects of splicing defects on the production of intron-contained snoRNAs. Will also include a brief update with unpublished preliminary data. The vast majority of this work was completed by Yanriu Liu and Samuel DeMario. My contribution to this work was the RNA FISH Microscopy and the NanoPore sequencing in the update with computational assistance from Kevin He.....**Pages: 72-85**

## List of Figures

Due to the mixed publication status of the work being presented in this thesis, figures will be globally numbered in this work in the following system: number1.number2.letter, where number 1 will represent the chapter of the figure, number 2 will represent the appearance of the figure in said chapter, and the letter will represent a panel within the figure. Supplemental figures will be specifically referenced as they are labeled.

<b>Figure 2.1:</b> Global analysis of 3'-end formation and splicing upon CPA or splicing inactivation.....	<b>Page: 18</b>
<b>Figure 2.2:</b> Impact of 3'-end processing defects on splicing efficiency.....	<b>Page: 21</b>
<b>Figure 2.3:</b> Impact of upstream readthrough transcripts and 5' extensions on splicing efficiency.....	<b>Page: 24</b>
<b>Figure 2.4:</b> Activation of novel Splice sites promoted by deficient transcription termination.....	<b>Page: 27</b>
<b>Figure 2.5:</b> Intergenic splicing events detected upon inactivation of CPA factors.....	<b>Page: 29</b>
<b>Figure 2.6:</b> Diagrams of the PDC1-RPL36A Fusion Constructs.....	<b>Page: 40</b>
<b>Figure 3.1:</b> Comparison of mRNA levels before and after nuclear depletion of CPA endonuclease Ysh1p.....	<b>Page: 51</b>
<b>Figure 3.2:</b> Some mRNA primary 3' ends are dependent on the presence of Reb1p.....	<b>Page: 52</b>
<b>Figure 3.3</b> Sequence analysis reveals enrichment for adenosines genomically encoded directly downstream of CPA independent mRNAs.....	<b>Page: 53</b>
<b>Figure 3.4:</b> Genomically encoded poly(A) tracts are transcribed as pseudo poly(A) tails.....	<b>Page: 54</b>
<b>Supplemental Figure 3.3.1:</b> CPA Independent RNAs are not particularly stable.....	<b>Page: 55</b>
<b>Supplemental Figure 3.3.2:</b> Other classes of mRNA transcriptional roadblocks termination.....	<b>Page: 56</b>

<b>Figure 4.1:</b> Rnt1p protein levels and localization remain unchanged during and after salt stress.....	<b>Page: 61</b>
<b>Figure 4.2:</b> Analysis of <i>BDF2</i> RMD in various stress conditions.....	<b>Page: 63</b>
<b>Figure 4.3:</b> Anchoring away mRNA biogenesis or mRNA export factors triggers RMD hyperactivation of <i>BDF2</i> in absence of stress.....	<b>Page: 65</b>
<b>Figure 4.4:</b> The identity of the Rnt1p cleavage site (RCS) influences RMD hyperactivation in stress conditions.....	<b>Page: 67</b>
<b>Figure 5.1:</b> Hybrid forms of intron-encoded snoRNAs are produced upon splicing inactivation.....	<b>Page: 75</b>
<b>Figure 5.2:</b> Detection of hmsnoRNAs by direct single-molecule Oxford Nanopore RNA sequencing.....	<b>Page: 76</b>
<b>Figure 5.3:</b> hmsnoRNAs acquire their 3' ends through processing by the nuclear exosome but are mostly unaffected by cytoplasmic quality control pathways.....	<b>Page: 77</b>
<b>Figure 5.4:</b> hmsnoRNAs generated by SS mutations are degraded by Xrn1p and Rat1p and share some of the features of mature snoRNPs.....	<b>Page: 78</b>
<b>Figure 5.5.</b> hmsnoRNAs are detectable by NanoPore sequencing.....	<b>Page: 85</b>

## ACKNOWLEDGEMENTS

In completion of this work I would like to acknowledge and thank the following for their contributions to this work academically, emotionally, and with their time: Kevin He, Andreas J Krumbein, Dean Neutel. To the above and the rest of the Chanfreau lab who may not have worked directly on the research presented here, but were instrumental in my life and making coming to lab every day worth doing. And lastly, to my advisor, Guillaume Chanfreau, who I feel lucky to have chosen as my advisor all those years ago who has provided nothing but support as I navigated grad school and encouraged me to pursue my goals both in and out of the lab.

## Vita

Keaton Barr

### Education

#### **-Southern Oregon University, Ashland, OR**

- Chemistry with Honors-ACS Biochemistry focus B.S., French Minor, cum laude, 2017

#### **-University of California, Los Angeles**

- M.S. Biochemistry, Molecular, and Structural Biology, 2019
- PhD Candidacy in Biochemistry, Molecular, and Structural Biology, May 13, 2019
- PhD in Biochemistry, Molecular, and Structural Biology (in progress)

### Undergraduate Research Experience

#### **-Senior Capstone SOU under Dr. Steven Petrovic: Analysis of Pesticides in Rogue Valley Honey by QuEChERS, GCMS, and LCMS**

- Adapt and modify industry standard techniques for use in the instrument lab at SOU
- Create instrumental methods and protocol manuals for future students to follow in the continuation of the research

### Thesis Graduate Research

#### **-BMSB PhD Program under Dr. Guillaume Chanfreau: Exploration of the Mechanism and Interplay of RNA Processing Pathways**

- Use various molecular biology techniques such as Northern blots, Western blots, FISH Microscopy, RNA sequencing, Next-Gen Sequencing, Oxford NanoPore Sequencing, and RT-PCR to explore the existence of and mechanisms for novel pathways of yeast transcriptional processes

### Major Presentations

#### **-Southern Oregon Arts and Research Seminar 2016:**

- Comparison of Fatty Acid Composition of Free Range and Cage Raised Chicken Eggs by GC-MS

#### **-Southern Oregon Arts and Research Poster Session 2017:**

- Analysis of Pesticides in Rogue Valley Honey by QuEChERS, GCMS, and LCMS

#### **-Fire Ecology and Policy (Science Policy Group General Meeting):**

- An extensive overview of the ecology of forest fires and the policy around their management

#### **-RNA Society Meeting June 2019 Poster Session- Barr, K\*., Roy, K., Chanfreau, G.**

- The Mep3 gene is protected from anti-sense pervasive transcription by a Pol III transcriptional roadblock

#### **-Cold Spring Harbor Eukaryotic mRNA Processing 2023- Barr, K\*., He, Kevin., Chanfreau, G.**

- Transcriptional Roadblocks: A novel mechanism of 3'-end formation independent of CPA machinery

### Publications

Barr, K., Goldberg, A., et. al., 2020, Water in Los Angeles: Rethinking the Current Strategy, Journal of Science Policy and Governance, <https://doi.org/10.38126/JSPG170202>

Wang, C., Barr, K., et. al., 2021, Inhibition of mRNA export triggers RNase III-Mediated decay of the *BDF2* mRNA RNA. 2021 Dec;27(12):1545-1556. doi: 10.1261/rna.078880.121. Epub 2021 Sep 8. PMID: 34497070; PMCID: PMC8594472.

Liu Y, DeMario S, He K, Gibbs MR, Barr KW, Chanfreau GF. Splicing inactivation generates hybrid mRNA-snoRNA transcripts targeted by cytoplasmic RNA decay. Proc Natl Acad Sci U S A. 2022 Aug 2;119(31):e2202473119. doi: 10.1073/pnas.2202473119. Epub 2022 Jul 25. PMID: 35878033; PMCID: PMC9351541.

## **Teaching and Mentorship Experience**

### **-Southern Oregon University, Ashland, OR**

-Math Tutoring January 2015- June 2017

-Head Science Tutor April 2016- 2017

- Work with University staff to establish the Science branch of the SOU tutoring center
- Work with a team of other tutors to bring quality one-on-one assistance to other students in math, biology, physics, and chemistry.
- Assist university staff in assessing potential new science tutors
- Develop work schedules for each quarter
- Communicate tutoring events and developments in the growth of the Science tutoring center to Physics, Chemistry, and Biology faculty

Private Tutor for SOU McNair Program 2015-2016

- Privately tutor other students in 400 level biochemistry

Chemistry Department Undergraduate Teaching Assistant 2015-2017

- 200 level General chemistry lab assistant helping the professor manage the lab and train students in the basics of lab science
- 300 level Organic spectroscopy lab assistant running weekly study sessions to help students in their mastery of understanding spectroscopy data and solving structures
- 300 level Biochemistry lecture assistant running exam study sessions and act as a substitute when the professor was at a conference

### **-University of California, Los Angeles**

- Chem 154 Teaching Assistant teaching students more advanced molecular biology lab concepts, techniques, and writing skills
- Chem 153A Teaching Assistant teaching students the building blocks of biochemistry and the introduction to metabolism
- Chem 153B Teaching Assistant teaching students the basics of DNA, RNA, and the central dogma of biology
- Undergraduate lab research mentor for BISEP, CARE Fellows, CARE Scholars, and MSD Scholars programs training undergraduate students with no prior experience as they have progressed through multiple undergraduate research programs and have now been hired to continue solo projects in the lab post-graduation or continues to graduate programs.

## **Awards and Honors**

Eagle Scout Rank (Boy Scouts of America)

Southern Oregon University *cum laude*

SOU Honor Roll, and Dean's List Fall 2014, Spring 2015-Spring 2017

Coyner Graf Memorial Chemistry Departmental Scholarship (\$1000) 2016-2017 A.Y.

SOU Chemistry Departmental Service Award 2015-2016 A.Y.

CMB Departmental Training Grant Fellowship (\$24,816/year) 2018-2020 A.Y.s

Mike Jung Excellence in Teaching Award 2022

## Chapter 1: Introduction to RNA Processing Pathways

## 1.1: General Introduction

Ribonucleic acids (RNAs) are a class of biological molecules composed of polymers of ribonucleotides that are assembled through the linkage of their sugar-phosphate backbone by RNA Polymerase (RNAP) enzymes <sup>Reviewed in 1</sup>. These RNA molecules, while all composed of the same building blocks can vary greatly in their function in the cell and play roles in numerous vital cellular processes such as relaying the genetic information of the cell's DNA to the ribosome for translation (mRNA)<sup>2</sup>, aiding in the modification of other RNAs (snoRNAs)<sup>reviewed in 3</sup>, carrying amino acids to the ribosome for protein production (tRNAs)<sup>4</sup>, targeting other RNAs for degradation (miRNAs)<sup>reviewed in 5</sup>, and more. With their wide-ranging importance to the health of the cell, the proper production of RNAs and the regulation of their production are equally important. The work presented in this thesis thus aims to further the understanding of how some of the processes an RNA may go through during its production are themselves regulated, how these various RNA processing pathways may interact with one another during their function, if previously unknown pathways exist, and the potential fate of a given RNA when various steps of their production are perturbed.

## 1.2: Transcriptional Termination and 3' end processing of RNAs

During mRNA transcription, when the active RNAP (RNA Pol II in *S. cerevisiae*) nears the end of its current transcription unit, various signals within the C-terminal domain (CTD) of the RNAP<sup>6</sup> and the nascent transcript<sup>7,8</sup> will recruit the Cleavage and Polyadenylation (CPA) Complex. After recruitment of the CPA complex to the nascent pre-mRNA, an endonuclease within the complex (Ysh1p in yeast, CPSF73 in higher eukaryotes)<sup>9</sup> will cleave the nascent pre-mRNA at a specific sequence known as the poly(A) site, thus liberating the pre-mRNA from the active RNA Polymerase while simultaneously creating an open RNA end on both sides of the cleavage <sup>reviewed in 10</sup>. For the newly cut pre-mRNA, this exposed 3'-end is then utilized by the poly(A) polymerase (PAP, Pap1p in *S. cerevisiae*) in the CPA complex to add a poly(A)-tail to

the 3'-end of the nascent pre-mRNA<sup>11</sup>: thus, creating the mature 3'-end of the mRNA, which along with other RNA modifications, signals that the mRNA is ready for export out of the nucleus for translation while also providing protection against various RNA decay mechanisms in the cell. On the other end of the CPA cleavage site is an exposed 5'-end of an RNA still connected to the actively transcribing RNA Polymerase. This exposed 5'-end is then bound by a 5'-3' exonuclease (Rat1p in yeast), which will then translocate down the RNA degrading it along the way until it collides with the active RNAP: dislodging it and terminating transcription<sup>12</sup>. In eukaryotes, this process is believed to be the process that every mRNA, aside from mammalian histone mRNAs, undergoes for the maturation of the pre-mRNA 3'-end and termination of RNA Polymerase transcription. However, for non-coding RNAs (ncRNAs) such as some small nucleolar RNAs (snoRNAs) and many other aberrant ncRNAs other termination pathways exist. One such pathway is the *S.cerevisiae* *NRD1-NAB3-SEN1* (NNS) pathway. During NNS-termination, RNA sequences within the nascent RNA are bound by the proteins Nrd1p and Nab3p, which in turn recruit the RNA helicase Sen1p. Sen1p, then translocates down the nascent transcript chasing RNA Pol II until they collide, knocking RNA Pol II off the DNA in a “torpedo” like mechanism, which terminates transcription<sup>13,14</sup>. RNAs produced by this mechanism are then typically polyadenylated by the Trf4/Air2/Mrt4p Polyadenylation (TRAMP) complex, a protein complex which binds to RNAs and adds short poly(A) tails. In contrast to the stabilization and protection effect seen with the longer PAP dependent poly(A) tails of mRNA, these shorter TRAMP dependent poly(A) tails typically assist in targeting these RNAs for decay/processing by the exosome<sup>15</sup>. Additionally, this mechanism of transcriptional termination is generally stochastic in its exact termination site. However, work from our lab<sup>16</sup> has shown that should RNA Pol II run into any natural barriers on the DNA, such as a DNA Binding protein like the transcription factors Reb1p, Rap1, Abf1, or the RNA Polymerase III transcription factor TFIIIB, its activity can be paused at the site of the collision. This pausing of RNA Pol II for an NNS-dependent transcript gives Sen1p the ability to consistently torpedo RNA Pol II at the

roadblock site, thus creating a more uniform termination to the RNAs at that locus.

### 1.3: Spliceosomal RNA Splicing

While many mRNAs are transcribed as one continuous piece of RNA, some mRNAs (the percentage of which can vary drastically by organism) have regions known as introns interspersed throughout their coding sequence which need to be removed through the process of splicing in order to produce the final mature mRNA from these genes. Within unspliced pre-mRNA three well conserved sequences characterize the intron: the 5' splice site, branch point, and 3' splice site. These three distinct sequences are not only recognized specifically by various components within spliceosome, but also carry out the splicing reaction as described further in this chapter. Generally occurring co-transcriptionally, this highly conserved two-step chemical process is carried out by the spliceosome, a large ribonucleoprotein (RNP) complex composed by 5 small nuclear RNAs (U1, U2, U4, U5, and U6 snRNAs) and multiple dozens of proteins depending on the species. Prior to catalysis, the U1 and U2 snRNP complexes bind to the 5' splice site and branch point respectively through base pairing interactions of their snRNA components and their corresponding binding sites<sup>17,18,19</sup>. These in turn recruit the U4, U5, and U6 tri-snRNP. From this point, the pre-catalytic complex undergoes multiple rounds of reconfigurations assisted by proteins such as Prp28p, Brr2p, and Prp2p: adding and removing multiple components, while also restructuring the substrate RNA and tri-snRNP in order to form the active site and properly position the branch point and 5' exon in preparation for catalysis<sup>19</sup>, reviewed in 20-22. During the first step of splicing, an adenosine within the branch point sequence of the intron performs a nucleophilic attack on the 5' splice site. This liberates the 5' exon and forms an intron lariat-3' exon intermediate<sup>20-22</sup>. After the first step of splicing is complete, the complex undergoes another reconfiguration induced by Prp16p, among other proteins to prepare for the second step of splicing<sup>23,24</sup>. In the second chemical step of splicing, this newly liberated 5' exon then performs a nucleophilic attack onto the 3' splice site: thus, joining together the two exons in separation from the intron lariat completing the process of splicing for a given

junction<sup>20-22</sup>. After completion of the splicing reaction, the spliced mRNA, intron lariat, and spliceosome components must be disassembled. First, the spliced mRNA is released from the complex by Prp22p<sup>25</sup>. Prp43p, among other proteins, then disassembles the remaining complex, recycling the components of the spliceosome and releasing the intron lariat<sup>26</sup>. Once released from the spliceosome, the intron lariat is then typically debranched by the debranching enzyme (Dbr1p in yeast) and degraded by exonucleases <sup>reviewed in 27</sup>. In addition to the primary function of splicing in the removal of introns from coding sequences of RNA, splicing, or rather alternative splicing, presents the cell with multiple mechanism through which it can modulate gene expression. Alternative splicing simply put is the process that an RNA undergoes when it is spliced in a pattern or with different splice sites to its normal configuration. The fate of an mRNA after alternative splicing can vary wildly. For example, RNAs in mammalian cells often contain many exons, each definitionally separated by an intron. Here, by changing the pattern of whether or not certain exons are included in the final spliced mRNA, the resulting proteins could be drastically different from one another; thus, allowing for the cell to create numerous protein isoforms from a singular gene sequence<sup>28</sup>. Alternative splicing can also occur where the 5' and/or 3' splice site utilized by the spliceosome are different, which would result in either the inclusion or removal of the RNA sequence between the primary and alternative splice sites<sup>28</sup>. Previous work has shown that mammalian cells can use alternative splicing to promote the inclusion of poison exons which trigger degradation by nonsense-mediated decay (NMD), a translation coupled decay pathway<sup>29,30</sup>. Previous work from the Chanfreau lab has demonstrated that in *S. cerevisiae*, alternative splicing events can shift the reading frame of the mRNA to introduce a premature stop codon (PTC). This inclusion of a PTC in the mRNA by alternative splicing was shown to then target these mRNAs for degradation by NMD as well; therefore, giving the cell another mechanism, through which it can regulate the levels of these mRNA species<sup>31</sup>.

## 1.4 mRNA Export

In order to be translated, the nascent, mature mRNA must be exported out of the nucleus to the cytoplasm so that it may interact with the ribosome. The recruitment of RNA export factor proteins to the mRNA occurs co-transcriptionally and is dependent on interactions between these export factors with cis-factors within the RNAP CTD<sup>32</sup> or other structures within the mRNA such as the poly(A)-tail or 5'-cap<sup>33</sup>. These RNA export factors then help guide the nascent mRNA to the nuclear pore complex (NPC), which is effectively a channel through the nuclear membrane into the cytoplasm. There at the NPC, the mRNA is then actively channeled through the pore in an ATP dependent manner, where it then is released into the cytoplasm to be transported to the ribosome for translation <sup>reviewed in 34</sup>. This primary function of mRNA export into the cytoplasm for translation through the NPC also creates another level through which the cell can modulate gene expression. That is, by retaining an mRNA species in the nucleus rather than exporting it, the ability of that mRNA to be translated and the other processes it may be subject to are different than if it were in the cytoplasm.

## 1.5 RNA Nucleases

In addition to the direct production and processing of RNAs, the cell must also have mechanisms through which it can both control the quality of said RNA products, but also break them down either to completion, or partially as a step within other pathways when necessary. This degradation and or processing is performed by nucleases: proteins responsible for cleaving nucleic acids. In the yeast *S.cerevisiae*, one mechanism that utilizes multiple types of RNA nucleases in the decay and or processing of its substrates is RNase III mediated decay (RMD). During RMD, the Rnt1 endonuclease binds to an AGNN tetraloop structure present at the top of double-stranded RNA (dsRNA) secondary structure and then cleaves the backbone of the dsRNA<sup>35</sup>. This cleavage of the RNA backbone by Rnt1p exposes 5' and 3' ends on the RNA, which can then act as substrates for exonucleases such as Rat1p a 5-3' exonuclease, and the

nuclear exosome, a large protein complex responsible for the 3'-5' degradation and processing of RNAs in eukaryotes. Within this nuclear exosome complex the majority of the proteins compose the core, which primarily acts a channel for RNAs positioning them for degradation by the two catalytic proteins in the complex: Dis3p and Rrp6p. In both cases, the active sites of these proteins cleave the backbone sequentially from the end through hydrolysis and release of nucleoside 5'-monophosphate molecules as they process their way through an RNA substrate

reviewed in 36,37.

## **1.6 RNA Production and Processing Pathways Interconnected**

These RNA production and processing pathways as well as others not discussed here however are not happening alone in a vacuum. Rather, these pathways are often working in consort to tightly manage the production of RNA in the cell, often directly interacting with one another: especially when the RNA binding proteins are in closely related pathways such as those involved in pre-mRNA maturation. One such example is the coupling of 3'-end processing and splicing in mammalian cells. As previously described, splicing and 3'-end processing of an mRNA both occur co-transcriptionally and are both required for the proper maturation of an mRNA. Research has shown that for these two processes, not only do their components directly interact as demonstrated in the binding of U2AF 65 (a protein in the human U2snRNP) and CF<sub>1m</sub> (a cleavage factor in the human CPA complex), but are also important for the proper function of one another<sup>38</sup>. Additionally, mutations in the poly(A) site sequence have been shown to cause defects in the splicing of mammalian terminal introns<sup>39</sup>. In addition to multiple RNA processing pathways functionalities being directly coupled together, in other cases, multiple RNA production and processing pathways can also come together in alternative ways to produce entire new classes of RNA in function of their normal purposes.

An example of this can be found in the production of snoRNAs. snoRNAs are a class of small non-coding RNAs that are bound by proteins to function in a snoRNP complex performing

the modification of other RNAs. These snoRNAs are produced in one of two ways in yeast: either as their own transcription unit, or within the intron of a coding gene. Transcription of the independently transcribed snoRNAs is terminated via NNS, after which a short poly(A) tail is added by the TRAMP complex before the 3' end is then trimmed by the nuclear exosome like most ncRNAs terminated by NNS and TRAMP. For these snoRNAs, nuclear exosome-mediated trimming is eventually blocked by the proteins of the snoRNP complex, thus protecting them from complete degradation. Additionally, for some of these independently transcribed snoRNAs, the 5'-end of these RNAs form the tetra-loop substrate structure for Rnt1p, which then cleaves the RNA creating an entry point for 5-3' exonucleases: allowing them to trim the snoRNA to the 5' RNP boundary. With trimming complete, the mature snoRNP complex is formed. For intron encoded snoRNAs however, they are transcribed as part of their host gene due to their location within the intron of said gene. In order to then be liberated from the intron the host mRNA must undergo splicing: therefore, releasing the intron lariat in which it resides. The intron lariat is then debranched and targeted for degradation by exonucleases as it normally would be. However, instead of fully degrading the debranched intron as a normal intron, the bound proteins of the snoRNP complex block again the exonuclease at the mature snoRNP boundary, leaving only the snoRNP reviewed in 40, 41.

These above examples clearly demonstrate that frequently, the direct coordination of multiple RNA processing pathways either through the interaction of their components or their chemical activities is crucial for the proper production or regulation for many RNA species. However, there is still much that is yet unknown in this regard for many pathways and their corresponding processes. The following work presented in this dissertation will discuss four explorations of this concept. The first two of which explore the relationship of CPA and splicing in *S. cerevisiae*, and the existence a novel CPA-independent 3' end processing pathway for mRNAs respectively. In the third work, we demonstrate a direct link between nuclear retention

of the *BDF2* mRNA and its degradation by RMD. Lastly, we evaluate how defects in splicing can affect the proper production of intron-encoded snoRNAs and how the cell handles the aberrant RNA species that is generated as a result of aforementioned defects.

## Chapter 1 References

1. Belogurov, G. A., & Artsimovitch, I. (2019). The Mechanisms of Substrate Selection, Catalysis, and Translocation by the Elongating RNA Polymerase. *Journal of molecular biology*, 431(20), 3975–4006.
2. BRENNER, S., JACOB, F., & MESELSON, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, 190, 576–581.
3. Kiss T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *The EMBO journal*, 20(14), 3617–3622.
4. HOAGLAND, M. B., STEPHENSON, M. L., SCOTT, J. F., HECHT, L. I., & ZAMECNIK, P. C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. *The Journal of biological chemistry*, 231(1), 241–257.
5. Shang, R., Lee, S., Senavirathne, G., & Lai, E. C. (2023). microRNAs in action: biogenesis, function and regulation. *Nature reviews. Genetics*, 24(12), 816–833.
6. Davidson, L., Muniz, L., & West, S. (2014). 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes & development*, 28(4), 342–356.
7. Hill, C. H., Boreikaitė, V., Kumar, A., Casañal, A., Kubík, P., Degliesposti, G., Maslen, S., Mariani, A., von Loeffelholz, O., Girbig, M., Skehel, M., & Passmore, L. A. (2019). Activation of the Endonuclease that Defines mRNA 3' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex. *Molecular cell*, 73(6), 1217–1231.e11.
8. Noble, C. G., Walker, P. A., Calder, L. J., & Taylor, I. A. (2004). Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA. *Nucleic acids research*, 32(11), 3364–3375.
9. Ryan, K., Calvo, O., & Manley, J. L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA (New York, N.Y.)*, 10(4), 565–573.
10. Girbig, M., Misiaszek, A. D., & Müller, C. W. (2022). Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nature reviews. Molecular cell biology*, 23(9), 603–622.
11. Tudek, A., Krawczyk, P. S., Mroczek, S., Tomecki, R., Turtola, M., Matylla-Kulińska, K., Jensen, T. H., & Dziembowski, A. (2021). Global view on the metabolism of RNA poly(A) tails in yeast *Saccharomyces cerevisiae*. *Nature communications*, 12(1), 4951
12. West, S., Gromak, N., & Proudfoot, N. J. (2004). Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage

sites. *Nature*, 432(7016), 522–525.

13. Steinmetz, E. J., Conrad, N. K., Brow, D. A., & Corden, J. L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature*, 413(6853), 327–331.
14. Carroll, K. L., Ghirlando, R., Ames, J. M., & Corden, J. L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA (New York, N. Y.)*, 13(3), 361–373.
15. Wyers, F., Rougemaille, M., Badis, G., Rousselle, J. C., Dufour, M. E., Boulay, J., Régnauld, B., Devaux, F., Namane, A., Séraphin, B., Libri, D., & 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.
16. Roy, K., Gabunilas, J., Gillespie, A., Ngo, D., & Chanfreau, G. F. (2016). Common genomic elements promote transcriptional and DNA replication roadblocks. *Genome research*, 26(10), 1363–1375.
17. Siliciano, P. G., & Guthrie, C. (1988). 5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes & development*, 2(10), 1258–1267.
18. van der Feltz, C., & Hoskins, A. A. (2019). Structural and functional modularity of the U2 snRNP in pre-mRNA splicing. *Critical reviews in biochemistry and molecular biology*, 54(5), 443–465.
19. Plaschka, C., Lin, P. C., Charenton, C., & Nagai, K. (2018). Prespliceosome structure provides insights into spliceosome assembly and regulation. *Nature*, 559(7714), 419–422.
20. Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology*, 3(7), a003707
21. Matera, A. G., & Wang, Z. (2014). A day in the life of the spliceosome. *Nature reviews. Molecular cell biology*, 15(2), 108–121.
22. Wilkinson, M. E., Charenton, C., & Nagai, K. (2020). RNA Splicing by the Spliceosome. *Annual review of biochemistry*, 89, 359–388.
23. Schwer, B., & Guthrie, C. (1991). PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature*, 349(6309), 494–499.
24. Wilkinson, M. E., Fica, S. M., Galej, W. P., & Nagai, K. (2021). Structural basis for conformational equilibrium of the catalytic spliceosome. *Molecular cell*, 81(7), 1439–1452.e9.

25. Schwer, B., & Gross, C. H. (1998). Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *The EMBO journal*, 17(7), 2086–2094.
26. Arenas, J. E., & Abelson, J. N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proceedings of the National Academy of Sciences of the United States of America*, 94(22), 11798–11802.
27. Mohanta, A., & Chakrabarti, K. (2021). Dbr1 functions in mRNA processing, intron turnover and human diseases. *Biochimie*, 180, 134–142.
28. Wright, C. J., Smith, C. W. J., & Jiggins, C. D. (2022). Alternative splicing as a source of phenotypic diversity. *Nature reviews. Genetics*, 23(11), 697–710.
29. Lewis, B. P., Green, R. E., & Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), 189–192.
30. Fair, B., Buen Abad Najar, C. F., Zhao, J., Lozano, S., Reilly, A., Mossian, G., Staley, J. P., Wang, J., & Li, Y. I. (2024). Global impact of unproductive splicing on human gene expression. *Nature genetics*, 56(9), 1851–1861.
31. Kawashima, T., Douglass, S., Gabunilas, J., Pellegrini, M., & Chanfreau, G. F. (2014). Widespread use of non-productive alternative splice sites in *Saccharomyces cerevisiae*. *PLoS genetics*, 10(4), e1004249.
32. MacKellar, A. L., & Greenleaf, A. L. (2011). Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. *The Journal of biological chemistry*, 286(42), 36385–36395.
33. Viphakone, N., Sudbery, I., Griffith, L., Heath, C. G., Sims, D., & Wilson, S. A. (2019). Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. *Molecular cell*, 75(2), 310–323.e8.
34. Köhler, A., & Hurt, E. (2007). Exporting RNA from the nucleus to the cytoplasm. *Nature reviews. Molecular cell biology*, 8(10), 761–773.
35. Chanfreau, G., Buckle, M., & Jacquier, A. (2000). Recognition of a conserved class of RNA tetraloops by *Saccharomyces cerevisiae* RNase III. *Proceedings of the National Academy of Sciences of the United States of America*, 97(7), 3142–3147.
36. Kilchert, C., Wittmann, S., & Vasiljeva, L. (2016). The regulation and functions of the nuclear RNA exosome complex. *Nature reviews. Molecular cell biology*, 17(4), 227–239.
37. Vanacova, S., & Stefl, R. (2007). The exosome and RNA quality control in the nucleus. *EMBO reports*, 8(7), 651–657.

38. Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S. Z., Keller, W., Antoniou, M., & Vagner, S. (2006). An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *The EMBO journal*, 25(20), 4854–4864.
39. Niwa, M., & Berget, S. M. (1991). Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes & development*, 5(11), 2086–2095.
40. Kufel, J., & Grzechnik, P. (2019). Small Nucleolar RNAs Tell a Different Tale. *Trends in genetics : TIG*, 35(2), 104–117.
41. Webster, S. F., & Ghalei, H. (2023). Maturation of small nucleolar RNAs: from production to function. *RNA biology*, 20(1), 715–736

## **Chapter 2: Transcription termination is critical for splicing efficiency and fidelity**

## 2.1: The Paper

### Transcription termination promotes splicing efficiency and fidelity in a compact genome

Keaton Barr<sup>1,2</sup>, Kevin L He<sup>1,2</sup>, Andreas J. Krumbein<sup>1</sup> and Guillaume F. Chanfreau<sup>1,3</sup>

1 Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California Los Angeles, Los Angeles CA 90095-1569.

2. Equal contribution.

3. Correspondence: guillom@ucla.edu

#### Abstract

Splicing of terminal introns is coupled to 3'-end processing by cleavage and polyadenylation (CPA) of mRNAs in mammalian genes. Whether this functional coupling is universally conserved across eukaryotes is unknown. Here we show using long read RNA sequencing in *S.cerevisiae* that splicing inactivation does not result in widespread CPA impairment and that inactivation of CPA does not lead to global splicing defects. However, 5'-extensions due to termination defects from upstream genes lead to splicing inhibition in a length-dependent manner. Additionally, for some extended RNAs resulting from failed termination, we observed decreased splicing fidelity resulting in novel intergenic and long-range intragenic splicing events. These results argue against a broad coupling of splicing to CPA in *S.cerevisiae* but show that efficient CPA-mediated transcription termination is critical for splicing fidelity and efficiency in a compact genome.

## Introduction

Synthesis of most eukaryotic mRNAs requires three main processing steps: capping of the 5'-end of the mRNA by addition of a 7-methyl guanosine(7-meG), removal of intronic sequences by splicing, and 3'-end formation by cleavage and polyadenylation (CPA). In addition to generating mRNA 3'-ends, CPA also promotes transcription termination by RNA polymerase II. The three major mRNA processing reactions occur co-transcriptionally, as factors involved are known to associate with RNA polymerase II and/or the nascent transcripts (reviewed in <sup>1</sup>). These three main processing reactions also influence each other. At the 5'-end of the mRNA, capping stimulates splicing of the first, cap-proximal intron (<sup>2-4</sup>; reviewed in <sup>5</sup>). This stimulation is conferred by the cap-binding complex (CBC), which binds the 7-meG cap and promotes recruitment of the U1 snRNP, resulting in enhanced recognition of the first, cap-proximal exon 5' splice sites (SS) by splicing factors <sup>2,3,6</sup>. At the 3'-end of the mRNA, splicing of the terminal intron and recognition of the CPA site are intimately connected (reviewed in <sup>1,7</sup>). Interestingly, this coupling of splicing and 3'-end processing can occur independently from transcription<sup>8,9</sup>, and several interactions have been described between 3'-end processing factors and splicing factors. At the genomic scale, perturbation of 3'-end processing has been shown to result in inhibition of splicing of terminal introns<sup>10</sup>.

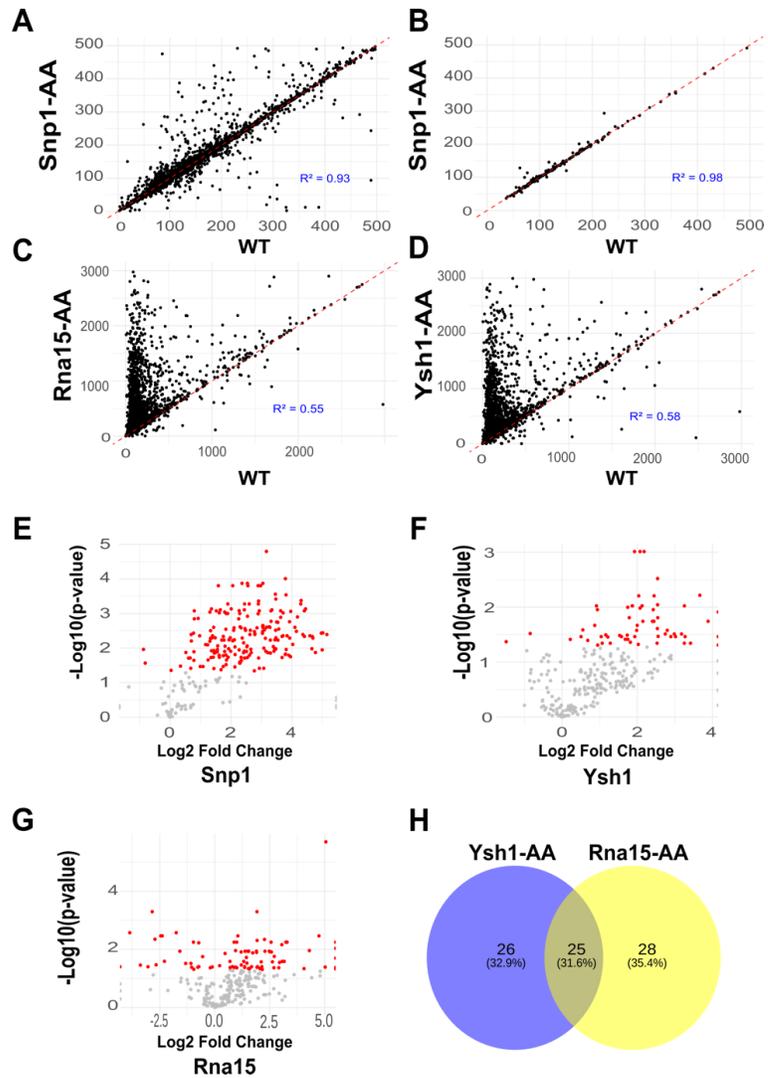
While there is broad evidence for reciprocal stimulation of splicing of terminal introns and 3'-end processing by CPA in mammalian genes, it is unclear whether this mechanism is universally conserved across eukaryotes. In support of this conservation, inactivation of the *S.pombe* homologue of splicing factor U2AF65 triggers 3'-end formation defects and transcriptional read-through<sup>11</sup>. In the yeast *S.cerevisiae*, most intron-containing genes contain only a single intron, which defines most introns as *de facto* 'terminal' and close to the cleavage and polyadenylation sites, especially for introns of ribosomal protein genes which constitute the quantitative majority of splicing substrates <sup>12</sup>. Some indirect evidence suggests that there might exist functional coupling between splicing and 3'-end formation in *S.cerevisiae*. For instance, the Ysh1 endonuclease involved in the cleavage reaction of CPA was first identified in a

genetic screen for splicing mutants<sup>13,14</sup>. In addition, physical and genetic interactions have been identified between splicing factors and the CPA machinery (reviewed in <sup>15</sup>). However no study has yet directly addressed the impact of inactivating CPA factors on splicing, or of inactivating splicing factors on 3'-end formation. In this study we address this question, using long-read RNA sequencing technology. Using rapid inactivation of splicing or 3'-end processing factors, we do not find evidence for broad reciprocal coupling between splicing and 3'-end processing. However, loss of termination promoting readthrough into intron-containing genes results in decreased splicing efficiency, likely because of the increased distance of the intron from the cap structure. In addition, loss of termination revealed dormant splice sites and resulted in the production of intergenic spliced mRNAs. These results underscore the importance of proper CPA-mediated termination to limit aberrant splicing events and enhance splicing efficiency.

**Inactivation of the U1 snRNP component Snp1 does not lead to prevalent 3'-end formation or termination defects.**

To analyze the reciprocal impact of splicing or 3'-end processing factors on splicing and polyadenylation, we assessed the impact of rapidly inactivating these factors on the transcriptome. Most splicing or 3'-end processing factors are essential in *S.cerevisiae*. To rapidly inactivate these RNA processing factors, we used the anchor away (AA) technique<sup>16</sup>, which promotes rapid export of a nuclear protein of choice out of the nucleus, resulting in their functional inactivation. This technique had previously been used to rapidly inactivate splicing<sup>17</sup> or 3'-end processing <sup>18,19</sup> factors in *S.cerevisiae*. We used AA strains of the 3'-end processing factors Ysh1 and Rna15, as well as the U1 snRNP component Snp1p, the yeast orthologue of the U1-70K protein. We chose to inactivate this early splicing factor to prevent engagement of transcripts in the splicing pathway and binding of proteins that may recognize splicing signals following E complex formation. Following treatment with Rapamycin to induce the export of tagged proteins out of the

nucleus, RNA processing efficiency was analyzed *in vivo* using long-read RNA sequencing. An untagged strain from the same genetic background (which alleviates any toxic effects of rapamycin treatment and downstream effects of gene expression) was treated for 2hrs with Rapamycin and used as a negative control (mentioned as WT in the remainder of this study). RNA samples were obtained for three replicates for each strain and sequenced by direct RNA sequencing using the Oxford Nanopore Technology (ONT) platform. We obtained an average of ca. 3 million reads per replicate (Table S1) and excellent reproducibility of RNA detection between replicates (Fig.S1). We then analyzed the impact of inactivation of these factors on 3'-end formation by measuring 3' UTR length. Analysis of global 3'-UTR lengths did not reveal any global impact of Snp1 inactivation on 3'-UTR length (Fig.1A).



**Figure 1. Global analysis of 3'-end formation and splicing upon CPA or splicing inactivation.**

**A.** Comparison of 3'UTR length between the Snp1-AA strain and the WT control for all *S.cerevisiae* genes detected by Nanopore sequencing. Numbers indicate the 3'UTR lengths in nucleosides.

**B.** Comparison of 3'UTR length between the Snp1-AA strain and the WT control for all *S.cerevisiae* intron-containing genes detected by Nanopore sequencing.

**C.** Comparison of 3'UTR length between the Rna15-AA and the WT control for all *S.cerevisiae* genes detected by Nanopore sequencing.

**D.** Same as (C) for the Ysh1-AA strain.

**E.** Volcano plot showing the ICGs exhibiting splicing defects in the Snp1-AA strain compared to WT. Red dots correspond to ICGs showing significant splicing changes.

**F.** Same as (E) for the Ysh1-AA strain.

**G.** Same as (E) for the Rna15-AA strain.

**H.** Venn Diagram showing the overlap of ICGs showing significant splicing defects in the Ysh1-AA and Rna15-AA strains.

This lack of effect could be due to the fact that intron-containing genes (ICG) represent only ~5% of genes in *S.cerevisiae*. We therefore analyzed 3'-UTR length only for ICG, and did not detect any significant changes in UTR length upon Snp1 inactivation (Fig.1B). This is in contrast to what was observed for Ysh1 or Rna15 inactivation, which resulted in a global increase in 3'-UTR size (Figure 1C), as expected since loss of these factors is known to induce termination defects.

Since Snp1 inactivation resulted in general splicing defects (see below), we conclude that splicing inactivation at an early stage of the splicing pathway does not result in major perturbations of 3'-end processing *in vivo*.

### **Inactivation of the cleavage and polyadenylation factors Ysh1 and Rna15 does not induce widespread splicing defects**

We next focused on analyzing the impact of inactivating Snp1, Ysh1 or Rna15 on pre-mRNA splicing. To this end, we quantitated unspliced reads in strains in which these factors were anchored away. Analysis of splicing efficiency in the Snp1-AA strain showed widespread increase of unspliced reads for many ICGs, as expected (Fig.1D). Some genes did not show significant accumulation of unspliced reads: either because their splicing is robust enough to resist to short-term inactivation of the U1 snRNP, or because unspliced

pre-mRNAs accumulation is limited by rapid degradation<sup>20,21</sup>. By contrast, inactivation of Rna15 or Ysh1 had a more modest effect on splicing (Fig1E, 1F). Some splicing defects overlap was observed upon inactivation of Rna15 or Ysh1(Fig.1G). However, inactivation of Ysh1 and Rna15 also resulted in unique set of genes for which splicing was affected (Fig.1G). This suggests that the impact of Ysh1 or Rna15 inactivation on splicing is not due to a general effect of CPA inactivation on splicing, but that each of these factors impact expression of the affected genes in a unique way which perturbs splicing indirectly, rather than through a general coupling mechanism between splicing and CPA.

### **Splicing defects are not prevalent in 3'-extended RNAs and unprocessed RNAs are subject to nuclear degradation**

To understand the mechanistic basis for the perturbation of splicing of genes impacted by Ysh1 or Rna15 inactivation, we first analyzed the splicing signals of ICGs showing a significant splicing defect in the Ysh1-AA or Rna15-AA samples. Sequence logos obtained for the 5'SS or branchpoint sequences of the RNAs with splicing defects in the Ysh1 or Rna15-AA strain were similar to those of all ICG (Fig.S2), suggesting that deficient splicing for these mRNAs was not due to intrinsically weaker splicing signals. To better understand the mechanism responsible for these splicing defects, we inspected ONT reads of ICGs to search for the presence of unspliced RNAs in transcripts with 3'-readthroughs in the Ysh1-AA or Rna15-AA samples. In general, intron-containing transcripts with long 3'-extensions following Ysh1 or Rna15 inactivation did not show increased accumulation of unspliced species. One example is shown in Fig.2A for the *RPL26A* gene, which exhibits frequent readthrough transcripts onto the neighbor gene *YLR345W*, none of which were unspliced. We calculated the splicing efficiency of RNA with 3'-extensions in the Ysh1-AA or Rna15-AA or control strain genome-wide, based on the percentage of unspliced reads, and compared this metric to the splicing efficiency of RNAs with no extensions. RNAs with 3'-extensions found

in the Ysh1-AA or Rna15-AA datasets did not show a large increase of unspliced reads compared to unextended transcripts found in the control wild-type sample or unextended transcripts detected in the Ysh1-AA or Rna15-AA samples (Fig.2B). Some exceptions include *RPL28*, for which intron retention was twice as abundant in 3'-extended transcripts compared to unextended transcripts in the Ysh1-AA or Rna15-AA (Fig.2C;

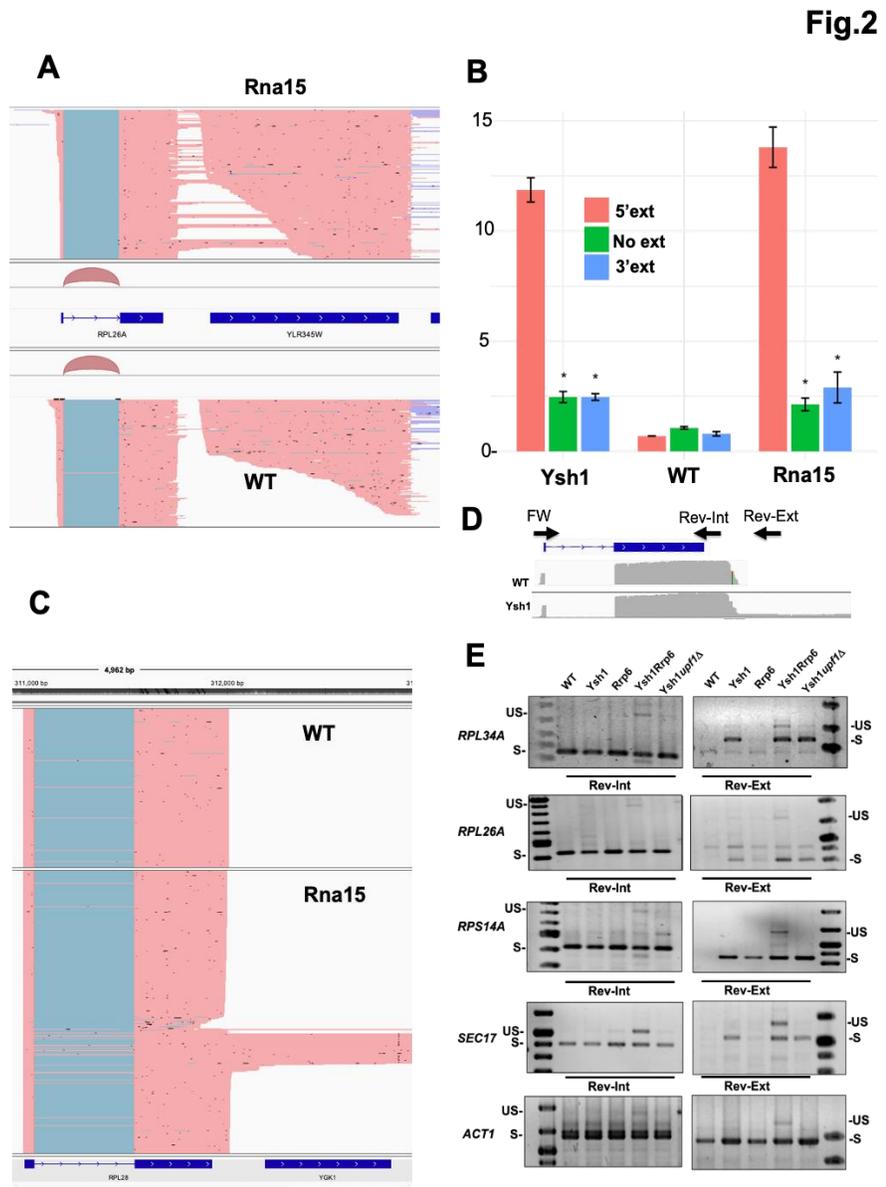


Fig.S3). Overall, these results argue that loss of recognition of the primary CPA or termination site does not impact the efficiency of splicing of upstream introns genome-wide.

**Figure 2. Impact of CPA inactivation on splicing efficiency.**

**A.** Selection of ONT reads from WT or Rna15AA in the *MRPL24-RPL26A-YLR345W* region showing the accumulation of 3' extended reads of *RPL26A* in the Rna15AA strain. Reads are colored in salmon; splicing event connecting reads are colored in blue. Sashimi plots showing the splicing of *RPL26A* are also included.

**B.** Bar graph showing the percentage of unspliced reads genome-wide as a function of 5' extension size in the Ysh1-AA, WT or Rna15-AA strains for RNAs with no extensions. The p-values obtained for the comparisons are: Rna15vs WT: No Ext: 0.005848; 3' extended: 0.003757;

Ysh1 vs WT: No Ext: 0.007634; 3' extended: 0.01342.

**C.** Selection of ONT reads from WT or Rna15-AA showing the accumulation of unspliced and 3' extended reads of *RPL28* in the Rna15AA strain.

**D.** RT-PCR scheme used to analyze the splicing status of normal and 3'-extended mRNAs. Shown is the Oxford Nanopore sequencing read coverage obtained for the WT and Ysh1-AA samples for one example gene (*RPS14A*) gene and approximate location of the primers used for the RT-PCR analysis. The Rev-Ext primer only anneal to 3'-extended forms detected upon Ysh1 or Rna15 inactivation. Similar schemes are used for the other genes.

**E.** RT-PCR analysis of splicing of normal and 3'-extended mRNAs. panels show RT-PCR products obtained for each of the genes indicated in the indicated strains. S=spliced; US = unspliced.

The ONT RNA sequencing platform relies on the presence of poly(A) tails for sequencing. For this reason, mRNAs that are 3' extended but unpolyadenylated were undetectable in the sequencing experiments described above. Therefore, we could not rule out that unpolyadenylated, 3'-extended readthrough transcripts might be predominantly unspliced when the CPA machinery is inactivated. In addition, we could not rule out that RNA degradation pathway such as the nuclear exosome, or nonsense-mediated decay might degrade unspliced transcripts, limiting their detection by Nanopore sequencing. To investigate the splicing status of unpolyadenylated, 3'-extended readthrough transcripts, we analyzed selected transcripts that did not show prevalent splicing defects in 3'-extended Nanopore reads by RT-PCR using total RNAs, which included unpolyadenylated RNAs. We used two different reverse primers: an internal reverse primer (Rev-Int) detects normal mRNAs, as well as 3'-extended ones; or a Rev-Ext primer that anneals only to 3'-extended RNAs and allowed the analysis of splicing for 3'extended forms (Fig.2D). Concomittant inactivation of the nuclear exosome component Rp6p by anchor away, or deletion of the gene encoding the NMD factor Upf1p was used to assess if unprocessed transcripts are subject to nuclear or cytoplasmic surveillance.

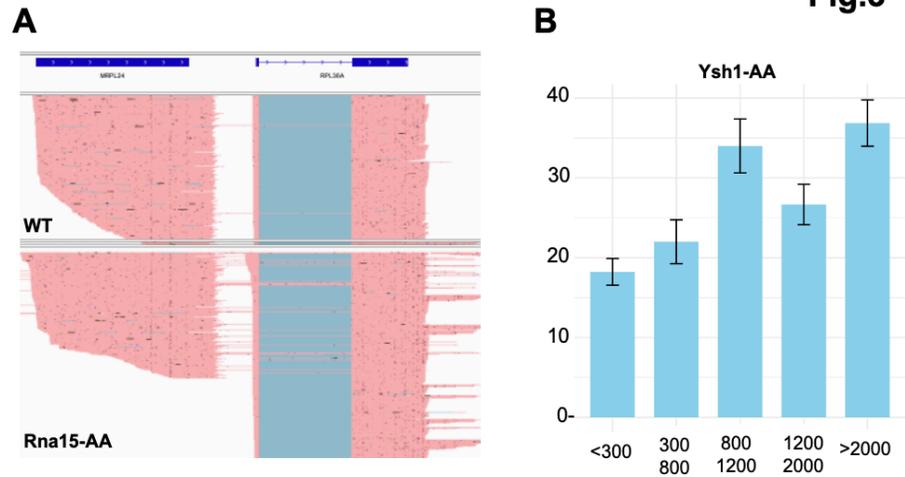
As show in Fig.2E, RT-PCR analysis using reverse primers that anneal only to extended RNAs showed that 3'-extended transcripts that were fully spliced were readily detected but absent in WT (or present at much

lower level for ACT1). This analysis did not reveal any major accumulation of unspliced 3'-extended transcripts in the Ysh1-AA strain. For all genes analyzed, a minor accumulation of unspliced 3'-extended transcripts could be detected upon inactivation of both Ysh1 and Rrp6,. This result shows that these 3'-extended and unspliced RNAs are efficiently degraded by the nuclear exosome. By contrast, inactivation of NMD did not result in an increase in unspliced RNAs detection, which showed that these transcripts are likely retained in the nucleus, perhaps due their lack of poly(A) tails. A similar minor accumulation of unspliced transcripts was observed with the Rev-Int primer in the Ysh1/Rrp6 double anchor away strain. We conclude that a fraction of unprocessed, 3'-extended RNAs fail to undergo splicing but they do not accumulate to detectable levels unless the nuclear exosome is inactivated. We note that concurrent inactivation of CPA factors and of Rrp6p has been shown to induce relocation of mRNAs to discrete nuclear compartments<sup>22</sup>, which may complicate the interpretation of these data.

### **Upstream termination defects result in length-dependent splicing defects for downstream intron-containing genes**

Inspection of the genes showing significant splicing defects in the Ysh1-AA or Rna15-AA samples revealed that unspliced reads were mainly found for RNAs with 5'-extensions resulting from readthrough transcription from upstream genes. These 5'-extensions are due to deficient termination from upstream transcription units, resulting in transcripts containing the upstream mRNAs, the intergenic region and the downstream ICG mRNA. One example is shown on Fig.3A where deficient termination of *MRPL24* transcription generated readthrough transcripts into the downstream ICG, leading to accumulation of unspliced *RPL36A* reads (Fig.3A). This effect was detected genome-wide, as shown by a large increase in the fraction of unspliced reads for reads containing 5'-extensions found in the Rna15-AA or Ysh1-AA strain relative to unextended transcripts found in any of the strains (Fig.2B). This effect is reminiscent to what

was observed upon nuclear depletion of the nuclear poly(A) binding protein Nab2<sup>23</sup>, which results in readthrough transcripts that are frequently unspliced.



**Figure 3. Impact of upstream readthrough transcripts and 5' extensions on splicing efficiency.**

**A.** Selection of ONT reads from WT or Rna15-AA in the *MRPL24-RP36A* region showing the accumulation of unspliced 5' extended reads of *RPL36A* in the Rna15-AA strain.

**B.** Bargraph showing the percentage of unspliced reads genome-wide as a function of 5' extension size in the Ysh1-AA strain. The p-values obtained for the comparisons are: <300 vs 800 -1200: 0.01486; <300 vs 1200-2000: 0.03149; <300 vs >2000: 0.009079.

**C.** Selected Nanopore reads from WT control cells in the *YEA4-GIM4* region. Reads at the top of the picture are readthrough transcripts from *YEA4* gene into *GIM4* which show a high fraction of unspliced reads.

**D.** Quantification of splicing defects for normal *GIM4* mRNAs (S), or readthrough mRNAs from the *YEA4* gene (L); p-value L vs S: 0.0003384. Numbers indicate the percentage of unspliced reads.

**E.** Selection of Nanopore reads from WT control cells in the *YDL012C* region.

**F.** Quantification of splicing defects for *YDL012C* mRNAs with short 5'exons (S), or with 5' exons longer than the annotated 5' exon (L); p-value L vs S: 0.01879. Numbers indicate the percentage of unspliced reads.

The results described above suggest that an increased distance from the 5'-end of the transcript to the intron is detrimental to splicing. To provide support for this hypothesis, we compared the length of 5' extensions found in transcripts showing significant splicing defect or no splicing defects. In the Ysh1-AA strain, the median 5' extension size for ICGs which showed a significant splicing defect was 1320 nt, while the median 5' extension size for ICGs which did not show a significant splicing defect was only 768 nt. In the Rna15-AA, the corresponding numbers were very similar, at 1426 nt and 833 nt, respectively. Binning reads obtained from the Ysh1-AA strain according to 5' extension lengths showed a threshold pattern, where transcripts with 5' extensions smaller than 800 nt showed a low level of intron retention, while transcripts with longer 5' extensions showed significantly stronger splicing defects (Fig.3D). A similar effect was observed for the Rna15-AA strain (Fig.S4). Overall, these data demonstrate that long 5' extensions that are caused by readthrough transcription upon CPA inactivation decrease splicing efficiency of downstream ICGs.

#### **mRNAs with naturally long 5'-extensions exhibit low splicing efficiency.**

The previous data showed that loss of termination due to nuclear depletion of CPA factors led to inefficiently spliced transcripts with 5'-extensions. To determine if this effect was solely due to the length of the 5'-exons/extensions or to indirect effects from CPA factors inactivation, we analyzed ONT data from WT cells to search for ICGs exhibiting reads with heterogeneous 5'-ends and analyzed their splicing efficiency as a function of 5'-exon length. The *GIM4* ICG is located just downstream of the *YEA4* gene, and we detected a high number of transcripts originating from *YEA4* and reading through *GIM4* in WT samples (Fig.3C). Strikingly, these readthrough transcripts showed a much lower splicing efficiency than the mRNAs originating from the bona fide *GIM4* promoter (Fig.3C,3D) as the level of unspliced reads for readthrough mRNA reached close to 50% (Fig.3D). For *YDL012C*, we detected some transcripts with heterogeneity in

5'-exon length, possibly due to usage of alternative transcription start sites (Fig.3E). As described above for *GIM4*, *YDL012C* mRNAs with naturally long 5'-exons were less efficiently spliced than those with short 5'-exons (Fig.3F), with an increase of ~5-fold in the amount of unspliced reads. We note that the numbers shown on this figure might be an underrepresentation of the actual splicing defects for these genes, as unspliced mRNAs are known to be targeted by various RNA surveillance systems<sup>21</sup>. Overall, these data demonstrate that long 5'-exons interfere with splicing efficiency. The most direct explanation for this effect is that long 5'-exon length increase the distance from the 5'-cap to the intronic sequences, decreasing the activation of splicing by the cap-binding complex<sup>5</sup>.

#### **Deficient transcription termination reveals dormant splice sites and promotes new splicing events**

We next analyzed the impact of inactivating Ysh1 or Rna15 on splicing fidelity, by searching for novel splicing events induced by inactivation of these factors. We hypothesized that RNA extensions caused by loss of termination upon CPA inactivation would expose sequences matching splice sites, which are otherwise not used in the context of normal termination. Inspection of the Ysh1-AA and Rna15-AA strain datasets identified several such cases. The 5' splice site of *RPS16B* was found to be spliced to a new 3'-splice site found within the *RPS16B* second exon (blue arrows, Fig 4A), in the Ysh1-AA or Rna15-AA strains, but not in WT. In addition, several reads showed frequent usage of a novel 5' splice in the middle of *RPS16B* exon2, spliced to the aforementioned 3' SS (red arrows, Fig.4A; Fig.S5). It is possible that the 3'SS found used in the Rna15 or Ysh1-AA strains is naturally used in a wild-type context. However, this would result in a very short transcript considering the length of the exons used, which is likely to be undetectable by Nanopore sequencing given the size limit for detection. In another case, we detected frequent usage of a novel 5'SS found at the end of exon2 of *RPL18A* spliced to the 3'SS of the following *RPS19A* gene upon Ysh1 or Rna15 inactivation (Fig.4B). This novel 5'SS showed a GUAAGU sequence which overlaps with the

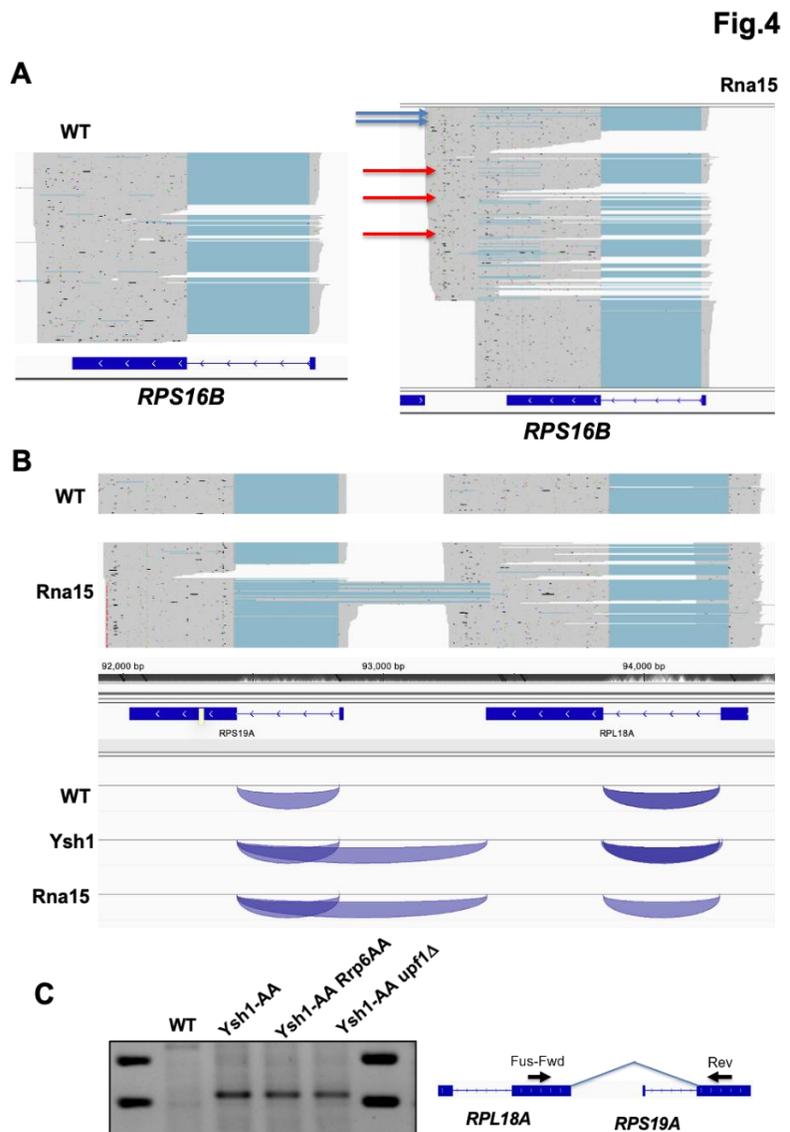
UAA stop codon of *RPL18A*.(Fig.S6). This splicing event resulted in the production of chimeric *RPL18A-RPS19A* transcripts, which corresponds to about 10% of the splicing events for extended species of *RPL18A* in Rna15 or Ysh1-AA. (Fig.4B).

**Figure 4. Activation of novel Splice sites promoted by deficient transcription termination.**

**A.** Selection of Nanopore sequencing reads for *RPS16B* in the WT control or Rna15AA strain. Blue arrows point to reads using the alternative 3'SS detected in the 3'UTR of *RPS16B*. Red arrows point to the isoform using the alternative 5'SS and 3'SS.

**B.** Selection of Nanopore sequencing reads for the *RPL18A* and *RPS19A* in the WT control or Rna15AA strain and Sashimi plots for WT, Rna15AA, and Ysh1AA showing the intergenic *RPL18A-RPS19A* splicing using the new 5'SS found at the end of *RPL18A*.

**C.** RT-PCR detection of the intergenic *RPL18A-RPS19A* splicing. RNAs extracted from WT control or Ysh1AA, Ysh1Rrp6AA, and Ysh1upf1 were analyzed by RT-PCR using the forward (Fwd) and reverse (Rev) primers indicated.



We validated the accumulation of this spliced product by RT-PCR in the Ysh1-AA (Fig.4C). Interestingly, this spliced isoform did not appear to be targeted by the nuclear exosome (Ysh1-Rrr6AA, Fig.4C), nor by nonsense-mediated decay (NMD) (Ysh1AA-upf1Δ, Fig.4C). The sequence of this chimeric mRNA would encode the Rpl18

protein in its entirety, fused to the protein sequence encoded by the second exon of *RPS19A*. The absence of premature stop codon explains the lack of effect of inactivating NMD. In the case of the duplicated *RPL18B* gene paralogue, the corresponding sequence is GUAAUC and we did not detect splicing between *RPL18B* and *RPS19B* at that site. The mismatches at positions 5 and 6 compared to canonical 5'SS sequences explains why such intergenic splicing events were not detected for the paralogue genes.

### **Deficient transcription termination promotes long-distance intergenic splicing events, including splicing-dormant genes.**

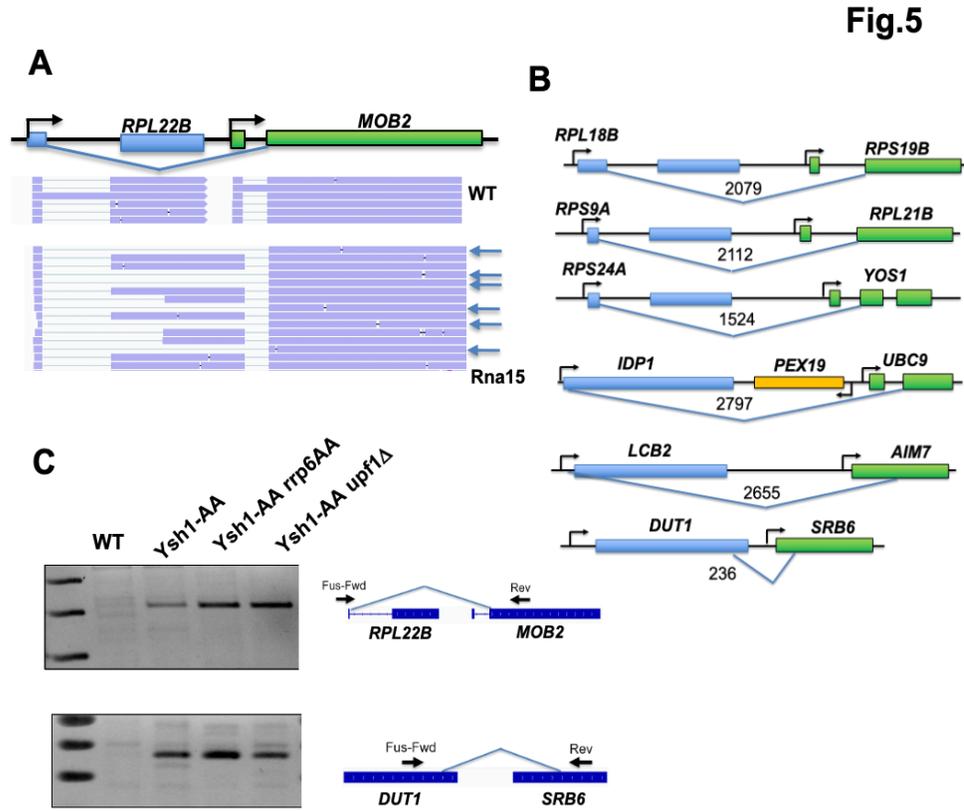
The previous example showed that loss of termination could promote the production of chimeric *RPL18A-RPS19A* transcripts. In *S.cerevisiae*, several intron containing genes (ICG) are located in tandem on the same strand (Table S2). Deficient termination due to inactivation of Ysh1 or Rna15 may therefore result in the production of transcripts covering multiple introns containing several splicing signals competing for usage, providing the opportunity for splicing events between tandem genes. We searched for such intergenic splicing events by identifying reads that showed splicing junctions between sequences belonging to different genomic features. After curating these junctions for overlapping transcription units, and for read alignment artefacts, we identified several cases of intergenic splicing triggered by loss of termination in the Ysh1-AA and/or Rna15-AA strains. The most frequent event detected was the splicing of the 5' SS of *RPL22B* to the 3' SS of the downstream *MOB2* gene (Fig.5A).

**Figure 5. Intergenic splicing events detected upon inactivation of CPA factors.**

**A.** *RPL22B-MOB2* genomic region and examples of Oxford Nanopore reads obtained from WT or RNA15-anchor away strains. Reads highlighted with blue arrows indicate intergenic *RPL22B-MOB2* splicing events.

**B.** Other examples of intergenic splicing events detected in the Ysh1-AA and/or Rna15-AA strains. Numbers indicate the distance between the 5' and 3' SS used.

**C.** RT-PCR analysis of intergenic splicing in WT, Ysh1-AA and derivative strains. The approximate locations of the forward (Fwd) and Reverse (Rev) primers used is indicated.



It was shown previously that an alternative intronic 5' SS is present in *RPL22B*<sup>24</sup>. This alternative 5' SS was also found to be spliced to the downstream 3' SS of *MOB2*. Reads showing splicing of exon1 of one intron-containing gene to the exon2 of a downstream ICG were also found for several other gene pairs: *RPL18B-RPS19B*, *RPS9A-RPL21B* and *RPS24A-YOS1* (Fig.5B). These events were found to occur upon inactivation of either Rna14 or Ysh1, showing that deficient transcription termination can promote intergenic splicing.

In addition to these events where natural splice sites were used, we also discovered cases where new splice sites are revealed. In the case of *IDP1*, a cryptic 5' SS is present at the beginning of the ORF, which becomes spliced to the 3' SS of the downstream *UBC9* gene (Fig.5B; Fig.S7). Strikingly, in the case of the *LCB2-AIM7* or *DUT1-SRB6* gene pairs, none of these genes are naturally spliced. However, Ysh1 or Rna15 inactivation generates readthrough species which results in the intergenic splicing between a 5' SS present

in the 5'UTR or ORF of the upstream gene (*LCB2* or *DUT1*) and a 3' SS in the ORF of the downstream gene (*AIM7* or *SRB6*, respectively; Fig.5B; Fig.S8). We used RT-PCR to validate two of these splicing events (Fig.5C), and test whether these intergenic spliced species are subject to nuclear surveillance or degradation by NMD. In agreement with the Nanopore sequencing data, intergenic splicing between *RPL22B* and *MOB2* or between *DUT1* and *SRB6* was detected only upon inactivation of Ysh1, but not in the WT control (Fig.5C). There was no major effect of inactivating Rrp6 or the NMD factor Upf1 on the accumulation of these intergenic spliced products, suggesting that these species are not a major target of these RNA degradation systems.

## Discussion

In this work we have studied the impact of rapid inactivation of splicing or 3'-end processing factors on splicing and 3'-end formation using long read ONT sequencing. The use of this platform allowed us to directly investigate the splicing efficiency and 3'-end of transcripts genome-wide. Strikingly, we did not find evidence for global reciprocal coupling between splicing and usage of downstream 3'-end processing signals. One of the limitations of this conclusion is that ONT sequencing relies on polyadenylated RNAs, and RNAs showing 3'-extensions in our sequencing data are still polyadenylated albeit at more distant sites. RT-PCR analysis of selected transcripts showed that a minor population of unspliced and 3'-extended transcripts can be detected in total RNAs (which include both polyadenylated and unpolyadenylated RNAs) if the nuclear exosome component Rrp6 is inactivated concomitantly with CPA factors (Fig.2E). While this may suggest some level of coupling of splicing to 3'-end processing, we note that the majority of these 3'-extended transcripts are still efficiently spliced (Fig.2E). In addition, simultaneous inactivation of CPA factors and of Rrp6p has been shown to induce relocation of mRNAs to discrete nuclear compartments<sup>22</sup>, which may complicate the interpretation of data involving these double

mutants. The involvement of the RNA exosome in degrading unprocessed RNAs is in agreement with previous studies showing that readthrough transcripts generated upon genetic inactivation of some CPA factors can be degraded by the exosome<sup>25</sup>. Overall our genome-wide analysis of the Rna15-AA or Ysh1-AA showed a global increase in 3'-UTR length (Fig.1); therefore we think that our genome-wide analysis still revealed the full extent processing defects in these strains, albeit at possibly lower levels than if all RNA surveillance pathways had been inactivated. Based on these experiments we conclude that there is no global coupling of splicing to 3'-end formation for most *S.cerevisiae* ICGs, although coupling might exist for a few transcripts such as *RPL28* (Fig.2). One reason for the loss of this coupling mechanism might be due to the fact that most transcription units in *S.cerevisiae* do not contain intronic sequences, unlike in mammalian genomes. Therefore the evolutionary pressure to maintain functional coupling between these two processes may have been lost if the signals that dictate the efficiency of each process are strong enough.

One of the major conclusions from our work is that loss of termination promoted by inactivation of CPA factors result in intron-containing transcripts with long 5'-extensions, which strongly decreased splicing efficiency. Strikingly, the size of the 5'-extensions negatively correlates with splicing efficiency. A negative impact on splicing for readthrough transcripts was reported previously upon nuclear depletion of the nuclear poly(A) binding protein Nab2<sup>23</sup>. While Nab2 is not a *bona fide* component of the CPA machinery, it is possible that the global impact of nuclear depletion of Nab2 on nuclear RNA decay<sup>26</sup> and/or on mRNA export<sup>27</sup> indirectly impairs transcription termination and/or CPA. Interestingly, we found that the deleterious impact of long 5'-extensions can be detected not only upon CPA inactivation, but also in a normal cells for transcripts exhibiting readthrough transcription from an upstream gene such as *GIM4* (Fig.3), or for mRNAs with heterogeneous transcription start sites such as *YDL012C*. These results shows that the splicing defects induced by long extensions are not intrinsically due to CPA deficiency but suggest

that an increase distance of the splicing signals to the 5'-cap structure is responsible for the impairment of splicing, as it would decrease the positive effect of the cap-binding complex on splicing. Alternatively, we cannot rule out that long 5'-extensions might promote secondary structures that impair the recognition of downstream splicing signals. However, given the general widespread effect of 5' extension length on splicing, we believe this hypothesis to be unlikely. Overall these data underscore the importance of proper termination to limit readthrough transcription onto downstream genes, which would largely inhibit splicing.

One of the most striking effect of inactivating CPA is the production of intergenic transcripts with multiple splicing signals, resulting in the detection of novel splicing events (Fig.4, 5). These results provide further evidence that inhibition of CPA does not impair splicing, as the extended species that are produced upon CPA inactivation are now spliced with novel patterns. Intergenic splicing events can be highly abundant in the case of *RPL22B-MOB2* or more rare, as they are sometimes detected with only a few reads in the Ysh1-AA or Rna15-AA samples. These spliced species might actually be underrepresented in our ONT sequencing datasets for two major reasons. First some of these species might be targeted by RNA surveillance pathways, even if we did not detect a major effect upon inactivating the nuclear exosome or NMD (Fig.4 and 5). For instance, exon skipped products (which functionally correspond to the intergenic spliced products we describe here) have been shown to be targeted by the exosome and Rat1 in *S.cerevisiae*<sup>28</sup>. In addition, some 5'-exons are very short, and while Nanopore sequencing offers good coverage of mRNAs, some 3'-end bias can still be observed in yeast<sup>29</sup>, which results in underrepresentation of transcript with short 5'-exons. We note that the distance involving some of these splicing events exceeds the size of the largest natural intron in *S.cerevisiae*, including several detected splicing events involving splice sites that separated than more than 2kb, including 2.6 and 2.8kb (Fig.5B). These results indicate that the yeast splicing machinery can operate at longer distances than previously thought. On the

other hand, readthrough between closely positioned transcription units can promote short distance splicing events such as the one detected between *DUT1* and *SRB6*. It is possible that some of these intergenic splicing events may occur in natural conditions and produce functional products if natural levels of transcriptional readthrough occurs, such as in stress conditions as observed in mammalian cells<sup>30</sup>. The compactness of the yeast genome and the close spatial positioning of transcription units make this more likely, and could potentially offer a way to produce proteins resulting from splicing of adjacent genes.

## **Acknowledgements**

We thank Douglas Black and members of the GFC lab for discussion and Alina Tong for comments on the manuscript. KH was supported by a Dissertation Year Fellowship from the UCLA Division of Graduate Education. Supported by NIGMS grant GM130370 to GFC.

## **Materials and Methods**

### **Yeast Strains and Growth.**

All yeast growth was performed in YPD media (1% w/v yeast extract, 2% w/v peptone, and 2% w/v dextrose). In all experiments cells were back diluted from an overnight solution to an O.D.600nm between 0.1 and 0.13 in 50mL per time-point and grown for approximately two doublings at 30°C, shaken at 200rpm. 50mL of cell culture would be harvested via centrifugation and flash frozen in liquid nitrogen for downstream use. Anchor away experiments were performed by replacing cell culture media of log-phase yeast with YPD+ 1µg/mL of rapamycin as described originally in <sup>16</sup>. Cells were then grown for an additional 2 hours in YPD+rapamycin before being harvested by centrifugation and flash frozen for downstream use. Strains used in this work can be found in Table S3.

### **RNA Extraction and Direct RNA sequencing.**

Total RNAs were extracted in triplicate via the standard phenol-chloroform technique <sup>31</sup> from anchor-away yeast strains. RNAs for the wild-type, ysh1-AA, and rna15-AA samples were then further processed according to the protocol described in the manufacturer's manual for ONT product (SQK-RNA004) for sequencing on the PromethION system (PRO-SEQ002) using RNA Flow Cells (FLO-PRO004RA). The only exception was the use of Maxima-H Minus as the RT enzyme as this was the recommended enzyme at the time of these experiments. RNAs for the snp1-AA samples were further processed per the manufacturer's manual for ONT product (SQK-RNA004) for sequencing on the MinION system (MIN-101B) using RNA Flow Cells (FLO-MIN004RA) using the current Induro RT enzyme protocol. RNA Flow Cells were washed according to the manufacturer's protocol for the Flow Cell Wash Kit (EXP-WSH00) in between sample runs.

### **RT-PCR Analysis**

RNAs for RT-PCR were generated from yeast cells as described above. RNA samples were then DNase I treated according to the manufacturer's instructions (Thermo Fisher EN0521). For cDNA synthesis from total RNAs the standard Maxima H Minus RT protocol (Thermo Fisher EP0751) was followed with gene specific primers and 100 units of enzyme per reaction. 1uL of each RT reaction was then used as template for a 20uL HiFi-PCR reaction using the corresponding gene specific primers (Supplemental Table S1). 5uL PCR products were then mixed with a 6x Bromophenol Blue and Xylene Cyanol loading dye and analyzed by agarose gel using 1.2-1.8% w/v agarose in 1xTAE (40mM Tris Base, 20mM acetic acid, 1mM EDTA, pH 8.2-8.4) depending on expected product size. Gels were then visualized by Ethidium Bromide staining on an Azure 300 Imaging System.

### **Bioinformatics and Statistical Analysis**

Reads were basecalled in real-time using the default algorithm for direct RNA sequencing on MinKNOW (version 6.2.6) with Dorado (version 7.6.7). Reads were then mapped to the *Saccharomyces cerevisiae* genome (S288C\_reference\_sequence\_R64-5-1) acquired from SGD using Minimap 2 (Version 2.17-r941). The following parameters were used during alignment “minimap2 -ax splice -G 3000” which set a maximum intron length of 3000 bp. Resulting bam files were coordinate sorted and converted to bed files using SAMtools<sup>32</sup> and Bedtools (<https://bedtools.readthedocs.io/en/latest/>) for downstream analyses. Reads were visualized using IGV (Version 2.12.3).

Reads were annotated with overlapping genes using bedtools intersect (v2.30.0). For read counts, we calculated the number of reads mapping to each gene using featureCounts as part of the Rsubread package (version 2.12.3) allowing for multiple overlaps. DESeq2<sup>33</sup> was used to quantify changes in gene expression and for normalizing counts to library size.

Counts were generated separately for whole genes and introns specifically. For a read to be counted as unspliced it needed to overlap an annotated intron by at least 6 bases. Ratios of unspliced were generated by dividing intron counts by the gene counts. For extension analysis, reads were classified into 4 groups. “3 prime extended” are reads which overlap the gene of interest but terminate more than 150 bp downstream of the annotated PAS. “5 prime extended” overlap the gene of interest, but begin more than 100 bp upstream of the TSS. If a read met both of these conditions, it was assigned as an “intrusive transcript”. All other reads fall into the “no extension” category. Experimentally derived UTR annotations<sup>34</sup> were used for the TSS and PAS locations.

All statistical tests and p-values reported were based on a welch's t-test. For the splicing volcano plots the p-value was adjusted further using the Benjamini-Hochberg method to control for False Discovery Rate.

### **Data availability**

Custom Nanopore processing code is available on GitHub: <https://github.com/kevinh97/Yeast-Transcription-termination>. The long-read RNA sequencing data reported in this paper are accessible in the NIH SRA database under the accession number PRJNA1229592.

## References

1. Shenasa, H. & Bentley, D. L. Pre-mRNA splicing and its cotranscriptional connections. *Trends in Genetics* 39, 672–685 (2023).
2. Colot, H. V., Stutz, F. & Rosbash, M. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev* 10, 1699–1708 (1996).
3. Lewis, J. D., Izaurralde, E., Jarmolowski, A., McGuigan, C. & Mattaj, L. W. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev* 10, 1683–1698 (1996).
4. Izaurralde, E. *et al.* A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* 78, 657–668 (1994).
5. Rambout, X. & Maquat, L. E. The nuclear cap-binding complex as choreographer of gene transcription and pre-mRNA processing. *Genes Dev* 34, 1113–1127 (2020).
6. Görnemann, J., Kotovic, K. M., Hujer, K. & Neugebauer, K. M. Cotranscriptional Spliceosome Assembly Occurs in a Stepwise Fashion and Requires the Cap Binding Complex. *Mol Cell* 19, 53–63 (2005).
7. Martinson, H. G. An active role for splicing in 3'-end formation. *Wiley Interdiscip Rev RNA* 2, 459–470 (2011).
8. Niwa, M. & Berget, S. M. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev* 5, 2086–2095 (1991).
9. Niwa, M., Rose, S. D. & Berget, S. M. In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev* 4, 1552–1559 (1990).
10. Movassat, M. *et al.* Coupling between alternative polyadenylation and alternative splicing is limited to terminal introns. *RNA Biol* 13, 646–655 (2016).
11. Herzel, L., Straube, K. & Neugebauer, K. M. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res* 28, 1008–1019 (2018).
12. Ares Jr., M., Grate, L. & Pauling, M. H. A handful of intron-containing genes produces the lion's share of yeast mRNA. *RNA* 5, 1138–1139 (1999).
13. Noble, S. M. & Guthrie, C. Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics* 143, 67–80 (1996).
14. Chanfreau, G., Noble, S. M. & Guthrie, C. Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). *Science* (1979) 274, 1511–1514 (1996).
15. Herzel, L., Ottoz, D. S. M., Alpert, T. & Neugebauer, K. M. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* 18, 637–650 (2017).

16. Haruki, H., Nishikawa, J. & Laemmli, U. K. The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. *Mol Cell* 31, 925–932 (2008).
17. Liu, Y. *et al.* Splicing inactivation generates hybrid mRNA-snoRNA transcripts targeted by cytoplasmic RNA decay. *Proc Natl Acad Sci U S A* 119, (2022).
18. Wang, C. *et al.* Stress-induced inhibition of mRNA export triggers RNase III-mediated decay of the BDF2 mRNA. *RNA* 27, 1545–1556 (2021).
19. Han, Z. *et al.* DNA-directed termination of RNA polymerase II transcription. *Mol Cell* 83, 3253–3267.e7 (2023).
20. Sayani, S., Janis, M., Lee, C. Y., Toesca, I. & Chanfreau, G. F. Widespread Impact of Nonsense-Mediated mRNA Decay on the Yeast Intronome. *Mol Cell* 31, 360–370 (2008).
21. Sayani, S. & Chanfreau, G. F. Sequential RNA degradation pathways provide a fail-safe mechanism to limit the accumulation of unspliced transcripts in *Saccharomyces cerevisiae*. *RNA* 18, 1563–72 (2012).
22. Carneiro, T. *et al.* Inactivation of cleavage factor I components Rna14p and Rna15p induces sequestration of small nucleolar ribonucleoproteins at discrete sites in the nucleus. *Mol Biol Cell* 19, 1499–1508 (2008).
23. Alpert, T., Straube, K., Carrillo Oesterreich, F. & Neugebauer, K. M. Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. *Cell Rep* 33, (2020).
24. Gabunilas, J. & Chanfreau, G. Splicing-Mediated Autoregulation Modulates Rpl22p Expression in *Saccharomyces cerevisiae*. *PLoS Genet* 12, e1005999 (2016).
25. Torchet, C. *et al.* Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs. *Mol Cell* 9, 1285–1296 (2002).
26. Schmid, M. *et al.* The Nuclear PolyA-Binding Protein Nab2p Is Essential for mRNA Production. *Cell Rep* 12, 128–139 (2015).
27. Green, D. M. *et al.* Nab2p is required for poly(A) RNA export in *Saccharomyces cerevisiae* and is regulated by arginine methylation via Hmt1p. *J Biol Chem* 277, 7752–7760 (2002).
28. Egecioglu, D. E., Kawashima, T. R. & Chanfreau, G. F. Quality control of MATa1 splicing and exon skipping by nuclear RNA degradation. *Nucleic Acids Res* 40, 1787–1796 (2012).
29. DeMario, S., Xu, K., He, K. & Chanfreau, G. NanoBlot: A Simple Tool for Visualization of RNA Isoform Usage From Third Generation RNA-sequencing Data. *bioRxiv* 2022.10.26.513894 (2022) doi:10.1101/2022.10.26.513894.
30. Vilborg, A., Passarelli, M. C., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Widespread Inducible Transcription Downstream of Human Genes. *Mol Cell* 59, 449–461 (2015).
31. Wang, C. *et al.* Rrp6 Moonlights in an RNA Exosome-Independent Manner to Promote Cell Survival and Gene Expression during Stress. *Cell Rep* 31, (2020).

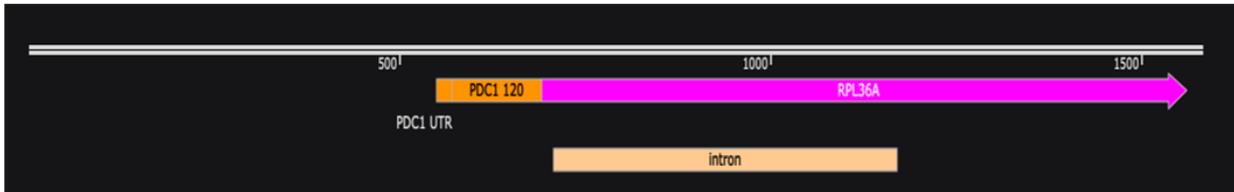
32. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).
33. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 1–21 (2014).
34. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science (1979)* 320, 1344–1349 (2008).

## 2.2 Ongoing work and future directions

Through the work described in chapter 2.1, we have demonstrated that unlike mammalian cells, the coupling of 3'-end processing and splicing is not generally observed in *S. cerevisiae*. We additionally provided some mechanistic insight towards improper splicing of 5' extended RNAs that agrees with previous literature. However, we wish to further explore the linkage between the length of the 5' extension and the efficiency of splicing of a given intron to support this work. To do so we have generated plasmids containing synthetic gene constructs of two different genes: ACT1, a traditional splicing reporter that does not gain a 5'-extension in CPA deficient environments; and RPL36A, an intron containing gene that does gain a 5'-extension in CPA deficient environments. In both cases we fused the gene of interest in-frame at the 5'-end to one of two different variants of the gene PDC1 downstream of its natural promoter region, which was chosen due to its natural high level of expression as well as its length of over 1000bp. In the first variant only the first 120 basepairs of PDC1 were included in the fusion: a length similar to that of the natural 5'UTR of ACT1, which should simulate the “expected” natural distance from the 5'-end of a given yeast mRNA to the 5'-splice site. The second construct however, contains the entirety of PDC1 sans stop codon fused directly to the first base of either ACT1 or RPL36A. This fusion then would simulate the presence of a long 5'-extension on the mRNA without the need to deplete the CPA machinery. Together, these synthetic genes will allow us to determine not only if the splicing defects *in vivo* due to the increased distance to the 5'-end from the 5'-splice site in these extended species, but also if the

absence of interactions between the CPA and splicing machinery is playing some role as well. Example diagrams of the RPL36A constructs are shown in Figure 2.6.

**A**



**B**



**Figure 2.6 Diagrams of the PDC1-RPL36A Fusion Constructs. A.** Fusion construct of the first 120bp of PDC1 and its 5' UTR direct attached to RPL36A. The relevant components are labeled within. **B.** Fusion construct of the full PDC1 gene and its 5' UTR to RPL36A. The relevant components are labeled within.

**Chapter 3: Novel pathways of mRNA3'-end formation: independence from CPA  
Machinery**

### 3.1: Abstract

With the exception of histone mRNAs, 3'-end processing of most eukaryotic mRNAs is mediated by the cleavage and polyadenylation (CPA) machinery which recognize various signals within the nascent pre-mRNA to induce cleavage and poly(A) tail addition. This cleavage of the nascent pre-mRNA by the CPA machinery also promotes the termination of RNA Polymerase II (Pol II) transcription. In addition to CPA, other molecular mechanisms have been shown to mediate Pol II transcription termination. For instance, the Chanfreau lab previously showed that the presence of proteins tightly bound to DNA can help mediate transcription termination for yeast ncRNA genes by the NNS-pathway through transcriptional roadblocks. In this study we assessed whether 3'-end formation of yeast mRNAs can occur independently from specific CPA factors. We performed NanoPore RNA sequencing of *S. cerevisiae* strains in which the Ysh1 or Rna15 components of the CPA machinery were rapidly exported to the cytoplasm. This experiment revealed several examples of yeast mRNAs that are insensitive to inactivation of Ysh1 or Rna15, suggesting that some specific transcripts do not require all components of the CPA machinery for mRNA 3'-end maturation. This lack of sensitivity to the loss of either Ysh1 or Rna15 was independently validated for a few specific transcripts via both Northern Blot or RT-qPCR. Interestingly, many transcripts insensitive to CPA factor inactivation are generated from genes with potential transcriptional roadblocks such as GTFs or tRNA genes near the mature mRNA 3'-end. We hypothesize that proteins tightly bound to proximal downstream regions result in transcriptional roadblocks, which terminate Pol II transcription of specific mRNA genes. The release of these pre-mRNAs due to the termination of Pol II would generate the substrate for the poly(A) polymerase, producing mature polyadenylated transcripts without cleavage. Additionally, NanoPore sequencing data revealed variations at the 3'-end of transcripts in response to CPA inactivation: some mRNAs maintain their approximate 3'-ends, while others produce a slightly extended versions after CPA inactivation. Furthermore, for many

of these CPA independent genes, their primary or extended 3'-end correspond to genomically encoded poly(A) tracts, which we propose may not only assist in inducing Pol II termination, but also act as either a primer for Pap1p, or are sufficient stand-ins for a traditional poly(A) tail. These data suggest that mRNA roadblock and poly(A) tract termination may serve both as a primary termination mechanism for some genes, but also as a fail-safe mechanism for other genes. Overall, these studies identify a novel mechanism of mRNA 3'-end formation independent of cleavage factors.

### **3.2: Introduction**

The process of cleavage and polyadenylation (CPA) is the definitional pre-mRNA 3' end processing pathway in eukaryotes and also provides the mechanism through which Pol II is able to terminate its activity during mRNA transcription <sup>reviewed in 1,2</sup>. Firstly, as the actively transcribing Pol II nears the end of a given mRNA transcript, the proteins of the CPA complex are recruited to the active Pol II and nascent pre-mRNA during transcription by signals with the C-terminal domain of Pol II<sup>3</sup>. In the yeast *S. cerevisiae*, this CPA complex is composed of three subcomplexes: cleavage factor (CF) IA, a complex composed of Rna14p, Rna15p, Clp1, and Pcf11; CF IB, which is the singular protein Hrp1p; and CPF, a complex of 14 proteins in *S. cerevisiae*, importantly containing the active endonuclease, Ysh1p, and the poly(A) polymerase, Pap1p, of the greater CPA complex<sup>4</sup>. In general, the proteins of CF IA, through interaction of its Rna14p and Rna15p components<sup>5</sup>, and CF IB bind to U-rich sequences within the nascent pre-mRNA both upstream and downstream of the prospective cut-site and assist in the positioning of CPF through various protein interactions. Meanwhile, proteins such as Yth1p within the CPF complex will bind the poly(A) site, an A-rich sequence in the nascent pre-mRNA in yeast that, unlike its mammalian equivalent, is generally poorly conserved <sup>4, reviewed in 6</sup>. Once bound, these protein complexes restructure the RNA:CPA complex, preparing it for enzymatic activity<sup>4</sup>. Here, the endonuclease of the complex, Ysh1p<sup>7</sup>, will cleave the RNA backbone, liberating the nascent

pre-mRNA from Pol II and creating an exposed 3' or 5' end on either side of the cut. Upstream of the cut, this cleavage creates the loose 3' end that will act as the entry point for the Pap1p, which then adds a long poly(A) tail in the dozens to over 100 bases long depending on the transcript<sup>8</sup>, reviewed in<sup>9</sup>. The other side of the cleavage site however, provides a mechanism through which to terminate the active Pol II transcription. This exposed 5' end of the nascent RNA is the entry point for a 5' to 3' exonuclease Rat1p, which proceeds to degrade the RNA still being actively transcribed until it runs into Pol II: knocking Pol II from the DNA and terminating transcription<sup>10</sup>. In this torpedo mechanism, the induction of Pol II pausing at specific genomic sequences such as poly(T) tracts has been shown to assist in the efficient termination of Pol II transcription by destabilizing the Pol II/DNA complex; thus, making it easier to dislodge<sup>11</sup>. This pathway is generally considered to be the pathway through which virtually all eukaryotic mRNAs gain their mature 3' end and terminate transcription. However, while the CPA dependent pathway may be the general mechanism for mRNA 3' end processing and transcriptional termination, previously published work has also demonstrated the existence of a roadblock dependent mechanism, which through Pol II collisions with DNA bound proteins, is able to act as a fail-safe for CPA and also terminate pervasive transcription events<sup>12,13</sup>. These findings raise the question as to whether these transcriptional roadblocks may also act as a primary mechanism of mRNA 3' end production, as well as if novel, yet undiscovered transcriptional termination pathways exist. In this work, we analyze the effects of depleting the nucleus of various components of the CPA machinery and known roadblock proteins in *S. cerevisiae*. We show while the majority mRNA species are downregulated in the absence of CPA function, a large number of mRNAs are still stably produced with their polyadenylated native 3' ends, or still stable, polyadenylated, alternative 3' ends: indicating the existence of alternative 3' end processing pathway for these specific mRNAs. Additionally, we provide insight as to what some of these alternative mechanisms may be, such as transcriptional roadblocks or intrinsic sequences within the DNA.

### **3.3: Results**

#### **3.3.1: Nuclear depletion of CPA factors results in genome-wide loss of mRNA transcript levels for most but not all genes**

In order to identify mRNAs as potential subjects of alternative 3' end processing pathways we first set out to analyze the genome-wide effects of CPA deficiency. To do this we performed direct RNA NanoPore sequencing on RNAs from cells where various CPA proteins were depleted from the nucleus through the anchor away (AA) technique. During the AA experiment a tagged nuclear protein of interest is rapidly sequestered into the cytoplasm during rapamycin treatment due to the interaction of the tag and the protein FKBP12, which only occurs during the presence of rapamycin<sup>14</sup>. This allows for the study of nuclear processes in the inducible export of the aforementioned tagged protein out of the nucleus. We found that generally, the majority of mRNA transcripts detected by NanoPore sequencing decrease in amount after depletion of CPA proteins, however there is also a population of genes whose mRNA levels remain relatively stable or even increase instead (Fig 3.1.A). Of note, these sequencing datasets generated from total RNAs after 2 hours of CPA depletion with no selection of nascent RNA populations post treatment. As such, there is the possibility that some effects seen in this study are due to hyper-stable mRNAs, which are surviving the 2-hour treatment rather than being RNAs that are stably produced post CPA shut-off. To address this concern, we performed a half-life analysis of select examples by RT-qPCR and found that these mRNAs are not particularly stable when compared to CPA dependent mRNAs after Pol II shutoff (Supplemental Figure 3.3.1). These results are also validated by comparison to recently published 4sU-seq data<sup>11</sup>, which only sequences nascently transcribed RNAs through selective pull-down of 4-thiouracil labeled RNA (Figures 3.2.A and 3.2.B)<sup>15</sup>. From this genome-wide data, we then selected a subset of genes both “CPA independent” (Fig 3.1.B) and “CPA dependent” (Fig 3.1.C) and performed RT-qPCR, which confirmed the NanoPore results. Additionally, we

demonstrate that these phenotypes occur in multiple CPA deficient backgrounds (Fig 3.1.C and D). These data are generally consistent with results seen previously for CPA deficient backgrounds both genome-wide and upon direct comparison of data from similar works<sup>11</sup> at individual genes.

### **3.3.2: Some CPA independent mRNA 3' ends correspond with previously known transcriptional roadblocks**

With this list of CPA independent mRNAs, we then determined to look for key features that may provide insight as to a mechanism through which these mRNAs are producing their 3' ends and termination transcription. Transcriptional roadblocks by general transcription factors (GTFs) and RNA Polymerase III have previously been identified as being capable of inducing Pol II termination for ncRNAs<sup>15,16</sup> and acting as an independent fail-safe should CPA fail<sup>12</sup>. As such, we first looked for evidence if these roadblocks are also observed in our data for mRNA transcription. To do so, we analyzed the 3'-ends of select CPA independent RNAs in the previously described NanoPore sequencing of CPA-AA RNA samples, as well as previous 3'-end sequencing data of reb1p-AA datasets from our lab<sup>16</sup>, and previously published 4sU-seq of similar ysh1-AA RNAs<sup>11</sup> and Reb1p CHIP data<sup>17</sup>. We observe that for some mRNAs that have a Reb1p binding site directly downstream of their primary 3' end, these 3' ends are unaffected by the loss of CPA function. However, after the nuclear depletion of the GTF Reb1p, the primary 3' end of these mRNAs shifts downstream: indicating that the presence of Reb1p on the DNA is important for the formation of these specific mRNA's 3' ends (Figure 3.2.A and Figure 3.2.C). Alternatively, for some transcripts both the ysh1-AA and reb1p-AA show usage of the same alternative 3' end site and similar long-distance run-off in depletion of their corresponding protein. This suggests that for some genes, the CPA and roadblock mechanisms may work in concert for the production of their primary 3' ends (Figure 3.2.B and 3.2.C). We additionally identified multiple mRNAs which seem to generate CPA independent 3' ends through other

classes of transcriptional roadblocks such as tRNA gene transcription (Supplemental Figure 3.3.2).

### **3.3.3: Genomically encoded poly(A) sequences assist in CPA independent termination for some mRNAs**

While we did find many CPA independent 3' ends that correspond to the presence of nearby transcriptional roadblocks, many others did not. Previous work has demonstrated that specific sequences within the DNA are sufficient to induce Pol II pausing, which can help induce termination of transcription. We posited that perhaps a similar sequence dependent mechanism may also exist for some CPA independent mRNAs. As such, we collected and analyzed the sequences directly on either side of the 3' end of mRNAs detected after CPA depletion in the 3' end sequencing dataset. A sequence logo was then generated for these 3' end adjacent regions both globally, and for genes that had a fold-change above 0.8 when compared to their untreated equivalents. These sequence logos demonstrated that for RNAs that are relatively stable in the absence of Ysh1p, we observe an enrichment of adenosines directly following the 3' end of the transcript (Figure 3.3A). Additionally, we plotted both the relative frequency of a transcript to have a given number of adenosines as a function of the number of adenosines downstream of its 3' end in the populations of mRNAs with fold changes above and below 0.8 after Ysh1p depletion (Figure 3.3.B) as well as the fold change of a given mRNA as a function of the number of adenosines located after its 3' end (Figure 3.3.C). In both cases, longer genomically encoded poly(A) stretches are associated with higher fold change values. Together, these data suggest that genomically encoded poly(A) tracts may play a role in the proper termination for CPA independent mRNAs.

### **3.3.4: Genomically encoded poly(A) tracts are transcribed and represent the true 3' end of some CPA independent transcripts**

To determine the exact role these genomically encoded poly(A) may play in the CPA independent termination for these genes we began to explore specific examples of mRNAs that fell into this category. Strikingly, we found that the 3' end of many of these genes lands directly within this genomically encoded poly(A) stretch; therefore, meaning that the actual end of the transcribed mRNA was also a poly(A) sequence creating a pseudo poly(A) tail. For some genes, this alignment of poly(A) stretch and 3' end occurs directly at their primary 3' end (Figure 3.4.A). However, for other genes, the loss of CPA termination led to a stable 3' extension which then contained the poly(A) stretch (Figure 3.4.B). Due to the innate poly(A) based nature of the sequencing techniques utilized in this study, we were concerned that our data may have been skewed by internal poly(A) priming during the library preparation. As such, the presence of this pseudo poly(A) tail on the mRNA from these genomically encoded poly(A) tracts was then confirmed in a poly(A) independent manner through a modified 3' Race protocol (Figure 3.4.C). Here, the RNAs were *in vitro* GI-tailed, which created a unique non-poly(A) sequence at the 3' end to which an alternative RT primer for 3' RACE could bind.

### **3.4: Discussion**

Cleavage and polyadenylation is one of the hallmark steps on the pre-mRNAs journey to maturation and eventually translation. Additionally, it is generally considered to play a major role in, if not required for the termination of Pol II transcription for mRNAs<sup>1,2</sup>. In this work, we show that while the majority of mRNAs are downregulated, there exists a population of mRNAs in *S. cerevisiae* that are insensitive to the inactivation of multiple components of the CPA complex. Within this CPA insensitive mRNA population, a number of these mRNA 3' ends correspond to loci with known inducers of transcriptional roadblocks, which when subsequently depleted from the nucleus, leads to alterations of the corresponding mRNA 3' ends. These data directly

corroborate previous findings which also demonstrated wide-spread transcriptional roadblocks of Pol II as a mechanism for terminating transcription for both pervasive and coding genes by a range of roadblockers such as GTFs, tRNA loci, and centromeres<sup>13</sup>.

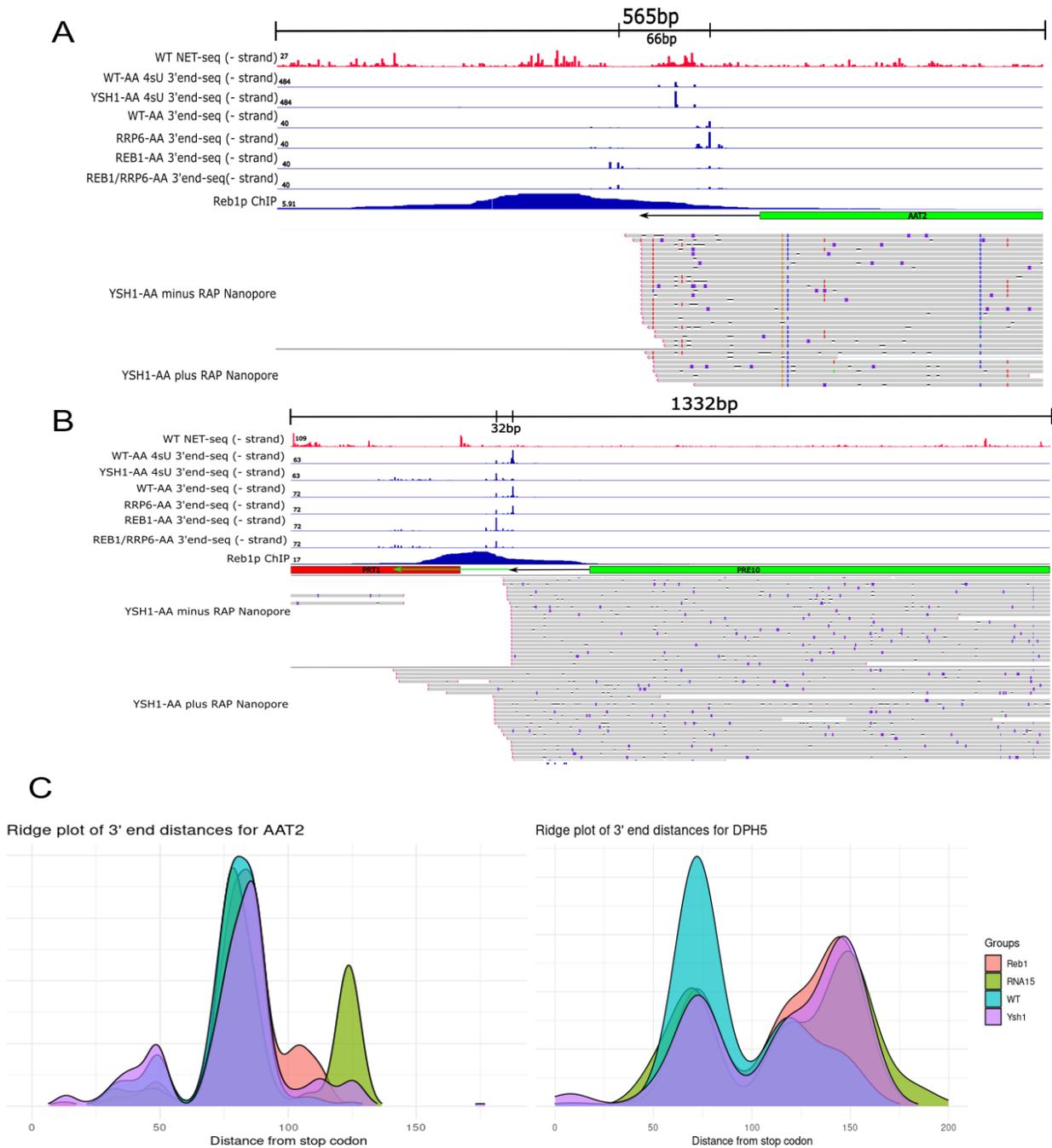
In addition to our confirmation of transcriptional roadblocks as an alternative mechanism through which Pol II can terminate, our findings demonstrate a sequence enrichment for poly(A) stretches genomically encoded in the DNA near the 3' end of these identified CPA independent transcripts. Additionally, for specific mRNAs studied here, these genomically encoded poly(A) tracts are transcribed and directly correlate to the 3' end of the transcript. We propose that these poly(A) sequences, frequently ranging from a dozen to sometimes exceeding 20 adenosines in length, represent an alternative transcriptional termination mechanism to CPA with a built-in stabilizing system for the mRNAs it generates: poly(A) tails. Homopolymer AT-rich sequences are known to induce pausing of Pol II and altogether have weaker interactions with the Pol II:DNA:RNA complex. As such, our identified poly(A) tracts may induce pausing and subsequent termination of transcription. However, due to the nature of these sequences, the resulting mRNA that is released contains a terminal poly(A) sequence of some length depending on the genomically encoded tract. These terminally transcribed poly(A)s may then act as a primer for processive addition by Pap1p, or be enough themselves for binding and subsequent protection by the yeast poly(A)-binding proteins Pab1p and Nab2p, which only require around 10-20 adenosines to bind<sup>18,19</sup>, thus, providing stabilization for these A-tract terminated mRNAs.

### **3.5 Future Directions**

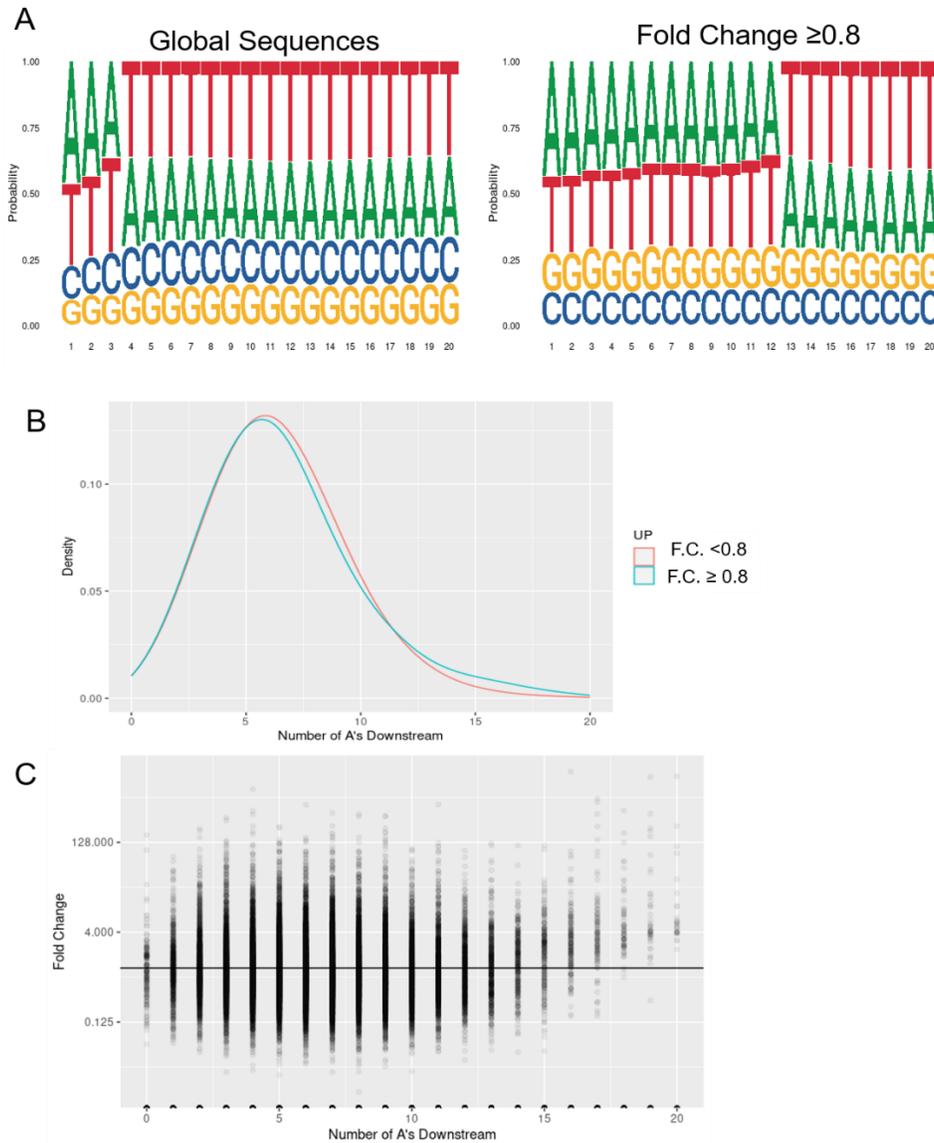
While the data presented here demonstrate an existing relationship between the presence of a genomically encoded poly(A) tail and independence from the CPA machinery, further experiments are needed to validate our proposed mechanism. In order to do this, I propose to use synthetic gene constructs similar to those being used in the ongoing work of chapter 2. These synthetic gene constructs would be constituted by one of two genes, one

being CPA dependent and the other being CPA independent with a genomically encoded poly(A) tail. From there, the 3' end of the CPA dependent gene would be replaced with a poly(A) tract, while the 3' end of the CPA independent gene would have its poly(A) tract removed. Should our hypothesis be correct, these swaps should either confer CPA independence or abolish it respectively, thus confirming if these poly(A) tracts constitute the proper requirements for CPA independent termination. One gene we particularly aim to study in this manner is that of *HSP12* (Figure 3.4.A), a heat-shock protein whose mRNAs terminate at a 12nt long poly(A) tract. Being a stress response gene, both gives innate biological relevance to *HSP12* as well as an easy means of induction. Additionally, we hypothesize that it may be beneficial for a stress response gene to have mechanisms of production that are more independent from the typical machinery as a means of functional insulation from specific stress responses.

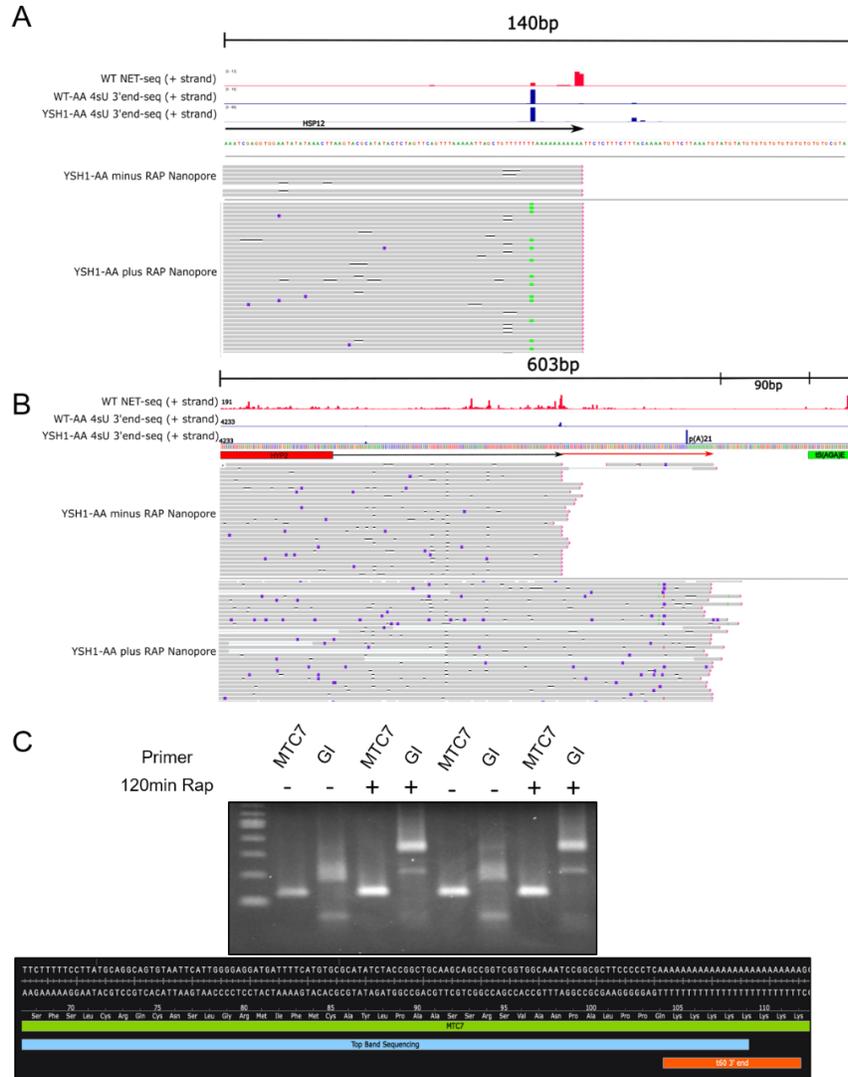




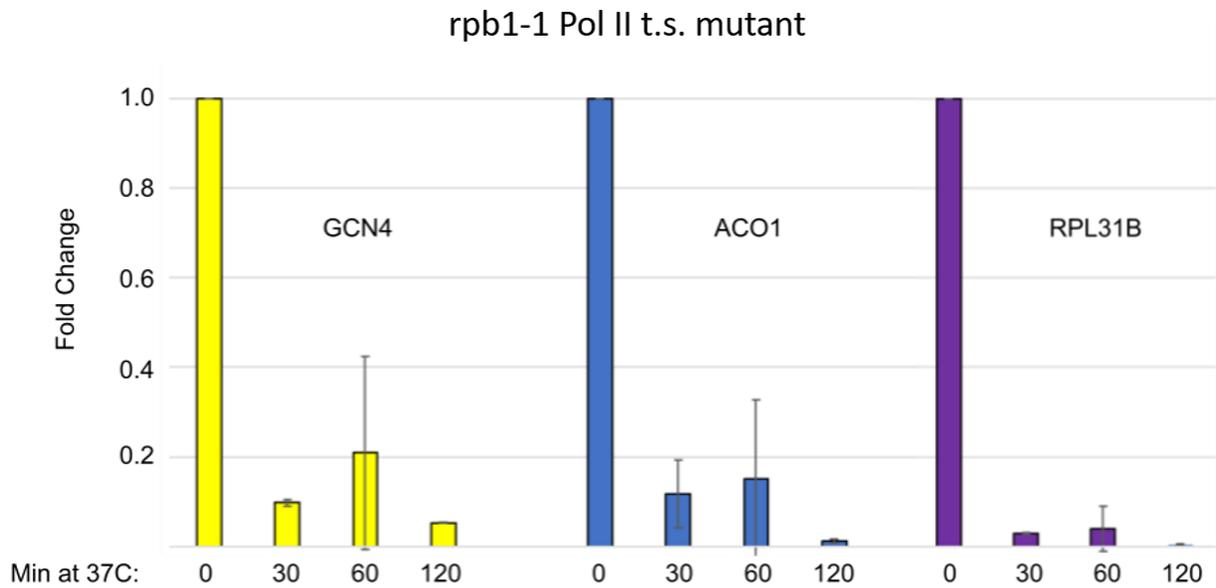
**Figure 3.2 Some mRNA primary 3' ends are dependent on the presence of Reb1p. A.** An array of Next-Gen and NanoPore sequencing, and Reb1p CHIP data at the AAT2 locus. **B.** An array of Next-Gen and NanoPore sequencing, and Reb1p CHIP data at the PRE10 locus. **C.** Ridge plots showing the distribution of 3' ends of AAT2 and DPH5 in WT, *ysh1-AA*, *rna15-AA*, and *reb1-AA*



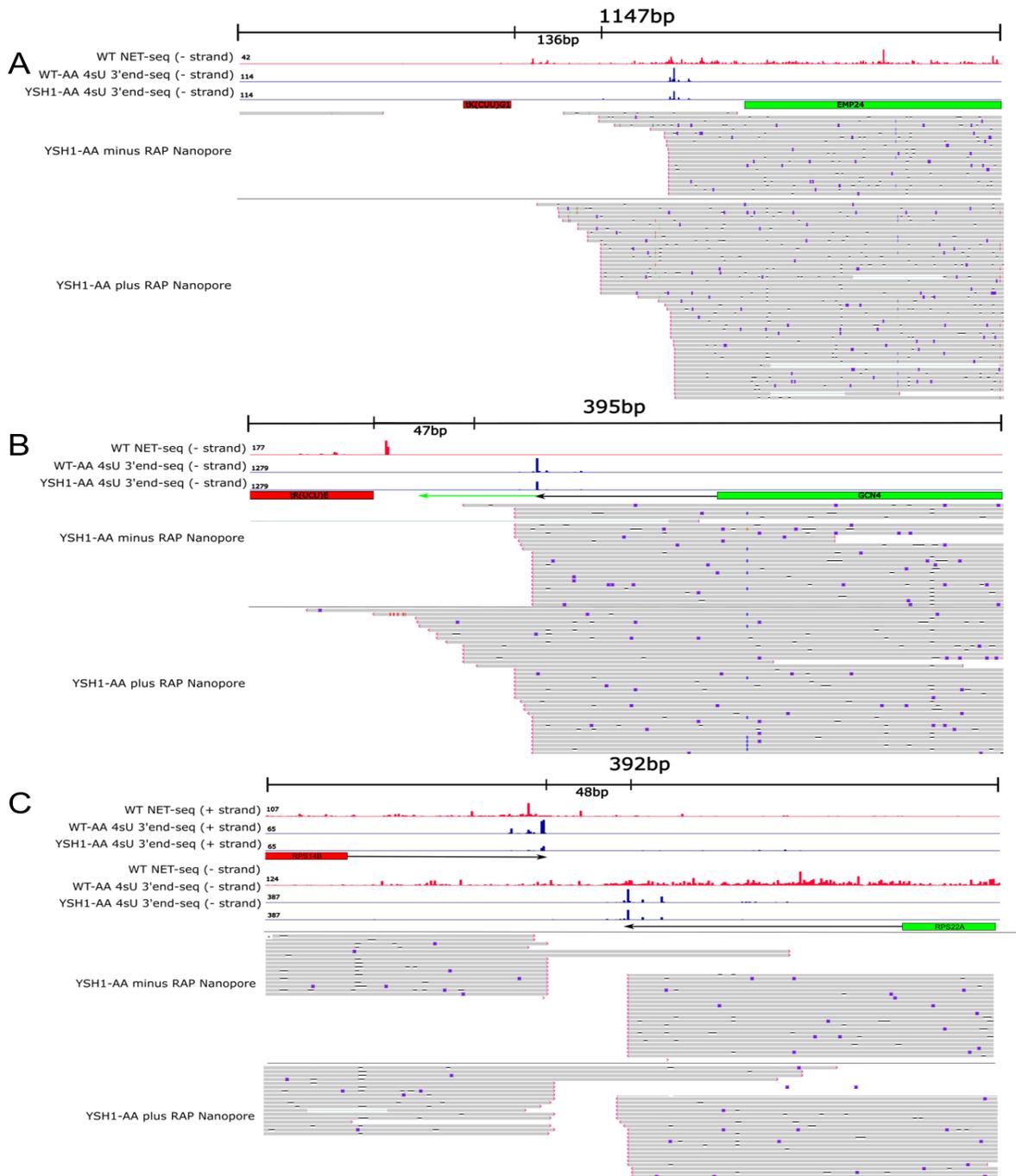
**Figure 3.3 Sequence analysis reveals enrichment for adenosines genomically encoded directly downstream of CPA independent mRNAs. A.** Sequence logos for the 20 bases directly after an RNA 3' end for all RNAs (left) and for RNAs with a Fold Change above 0.8 after ysh1-AA (right). **B.** Frequency of an RNA to have a given number of adenosines plotted for both RNAs below a fold change of 0.8 (red) and those above 0.8 (blue). **C.** Fold change after ysh1 AA plotted as a function of the number of As immediately downstream the 3' end.



**Figure 3.4 Genomically encoded poly(A) tracts are transcribed as pseudo poly(A) tails.** Next-Gen and NanoPore sequencing shown at the 3' end of the gene HSP12 in wild-type, and pre and post ysh1-AA samples. **B.** Next-Gen and Nanopore sequencing showing the 3' end and extension of MTC7 RNAs in wild-type, and pre and post ysh1-AA samples. **C.** Agarose gel of GI Race samples. RT was performed with either an internal primer for MTC7 or a primer for the GI tail. The top band of the GI tail Rap + lane was sequenced and aligned to MTC7 shown in blue. The 3' end identified by Nanopore sequencing is highlighted in orange.



**Supplemental Figure 3.3.1 CPA Independent RNAs are not particularly stable.** RT-qPCR of two CPA independent genes GCN4 and ACO1 and one CPA dependent gene RPL31B during a time course of Pol II shut off by the heat shock of Pol II t.s. mutant rpb1-1.



**Supplemental Figure 3.3.2: Other classes of mRNA transcriptional roadblocks**

**termination. A.** Next-Gen and Nanopore sequencing showing the 3' end of EMP24 labelled in green. The tDNA tK(CUU)G1 is labeled in red. **B.** Next-Gen and Nanopore sequencing showing the 3' end of GCN4 labelled in green. The tDNA tR(UCU)E is labeled in red. **C.** Next-Gen and Nanopore sequencing showing the 3' ends of RPS14B in red and RPS22A in green.

### Chapter 3 References:

1. Porrua, O., & Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nature reviews. Molecular cell biology*, 16(3), 190–202.
2. Girbig, M., Misiaszek, A. D., & Müller, C. W. (2022). Structural insights into nuclear transcription by eukaryotic DNA-dependent RNA polymerases. *Nature reviews. Molecular cell biology*, 23(9), 603–622.
3. Davidson, L., Muniz, L., & West, S. (2014). 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes & development*, 28(4), 342–356.
4. Hill, C. H., Boreikaitė, V., Kumar, A., Casañal, A., Kubík, P., Degliesposti, G., Maslen, S., Mariani, A., von Loeffelholz, O., Girbig, M., Skehel, M., & Passmore, L. A. (2019). Activation of the Endonuclease that Defines mRNA 3' Ends Requires Incorporation into an 8-Subunit Core Cleavage and Polyadenylation Factor Complex. *Molecular cell*, 73(6), 1217–1231.e11.
5. Noble, C. G., Walker, P. A., Calder, L. J., & Taylor, I. A. (2004). Rna14-Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA. *Nucleic acids research*, 32(11), 3364–3375.
6. Kumar, A., Clerici, M., Muckenfuss, L. M., Passmore, L. A., & Jinek, M. (2019). Mechanistic insights into mRNA 3'-end processing. *Current opinion in structural biology*, 59, 143–150.
7. Ryan, K., Calvo, O., & Manley, J. L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA (New York, N.Y.)*, 10(4), 565–573.
8. Tudek, A., Krawczyk, P. S., Mroczek, S., Tomecki, R., Turtola, M., Matylla-Kulińska, K., Jensen, T. H., & Dziembowski, A. (2021). Global view on the metabolism of RNA poly(A) tails in yeast *Saccharomyces cerevisiae*. *Nature communications*, 12(1), 4951
9. Rodríguez-Molina, J. B., & Turtola, M. (2023). Birth of a poly(A) tail: mechanisms and control of mRNA polyadenylation. *FEBS open bio*, 13(7), 1140–1153.
10. West, S., Gromak, N., & Proudfoot, N. J. (2004). Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016), 522–525.
11. Han, Z., Moore, G. A., Mitter, R., Lopez Martinez, D., Wan, L., Dirac Svejstrup, A. B., Rueda, D. S., & Svejstrup, J. Q. (2023). DNA-directed termination of RNA polymerase II transcription. *Molecular cell*, 83(18), 3253–3267.e7.
12. Colin, J., Candelli, T., Porrua, O., Boulay, J., Zhu, C., Lacroute, F., Steinmetz, L. M., & Libri, D. (2014). Roadblock termination by reb1p restricts cryptic and readthrough

- transcription. *Molecular cell*, 56(5), 667–680.
13. Colin, J., Candelli, T., Porrua, O., Boulay, J., Zhu, C., Lacroute, F., Steinmetz, L. M., & Libri, D. (2014). Roadblock termination by reb1p restricts cryptic and readthrough transcription. *Molecular cell*, 56(5), 667–680.
  14. Haruki, H., Nishikawa, J., & Laemmli, U. K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Molecular cell*, 31(6), 925–932.
  15. Rädle, B., Rutkowski, A. J., Ruzsics, Z., Friedel, C. C., Koszinowski, U. H., & Dölken, L. (2013). Metabolic labeling of newly transcribed RNA for high resolution gene expression profiling of RNA synthesis, processing and decay in cell culture. *Journal of visualized experiments : JoVE*, (78), 50195.
  16. Roy, K., Gabunilas, J., Gillespie, A., Ngo, D., & Chanfreau, G. F. (2016). Common genomic elements promote transcriptional and DNA replication roadblocks. *Genome research*, 26(10), 1363–1375.
  17. Harlen, K. M., Trotta, K. L., Smith, E. E., Mosaheb, M. M., Fuchs, S. M., & Churchman, L. S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell reports*, 15(10), 2147–2158.
  18. Brockmann, C., Soucek, S., Kuhlmann, S. I., Mills-Lujan, K., Kelly, S. M., Yang, J. C., Iglesias, N., Stutz, F., Corbett, A. H., Neuhaus, D., & Stewart, M. (2012). Structural basis for polyadenosine-RNA binding by Nab2 Zn fingers and its function in mRNA nuclear export. *Structure (London, England : 1993)*, 20(6), 1007–1018.
  19. Wigington, C. P., Williams, K. R., Meers, M. P., Bassell, G. J., & Corbett, A. H. (2014). Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions. *Wiley interdisciplinary reviews. RNA*, 5(5), 601–622

**Chapter 4: Stress-induced inhibition of mRNA export triggers RNase III-mediated decay of the *BDF2* mRNA**

# Stress-induced inhibition of mRNA export triggers RNase III-mediated decay of the *BDF2* mRNA

CHARLES WANG,<sup>1</sup> KEATON BARR,<sup>1</sup> DEAN NEUTEL,<sup>1</sup> KEVIN ROY,<sup>1,2</sup> YANRU LIU,<sup>1</sup> and GUILLAUME F. CHANFREAU<sup>1,2</sup>

<sup>1</sup>Department of Chemistry and Biochemistry, UCLA, Los Angeles, California 90095, USA

<sup>2</sup>Molecular Biology Institute, UCLA, Los Angeles, California 90095, USA

## ABSTRACT

The expression of bromodomain-containing proteins that regulate chromatin structure and accessibility must be tightly controlled to ensure the appropriate regulation of gene expression. In the yeast *S. cerevisiae*, Bromodomain Factor 2 (*BDF2*) expression is extensively regulated post-transcriptionally during stress by RNase III-mediated decay (RMD), which is triggered by cleavage of the *BDF2* mRNA in the nucleus by the RNase III homolog Rnt1p. Previous studies have shown that RMD-mediated down-regulation of *BDF2* is hyperactivated in osmotic stress conditions, yet the mechanisms driving the enhanced nuclear cleavage of *BDF2* RNA under these conditions remain unknown. Here, we show that RMD hyperactivation can be detected in multiple stress conditions that inhibit mRNA export, and that Rnt1p remains primarily localized in the nucleus during salt stress. We show that globally inhibiting mRNA nuclear export by anchoring away mRNA biogenesis or export factors out of the nucleus can recapitulate RMD hyperactivation in the absence of stress. RMD hyperactivation requires Rnt1p nuclear localization but does not depend on the *BDF2* gene endogenous promoter, and its efficiency is affected by the structure of the stem-loop cleaved by Rnt1p. Because multiple stress conditions have been shown to mediate global inhibition of mRNA export, our results suggest that the hyperactivation of RMD is primarily the result of the increased nuclear retention of the *BDF2* mRNA during stress.

**Keywords:** Rnt1p; RNase III; Bromodomain; mRNA export; stress

## INTRODUCTION

The covalent modification of histones is a major epigenetic mechanism that regulates the accessibility of DNA for damage repair, transcriptional activation or repression and heterochromatin formation (Lawrence et al. 2016). Histones acetylation is a well-studied covalent modification that is generally associated with transcriptional activation (Kurdستاني and Grunstein 2003). Bromodomain-containing proteins bind to acetylated histones and recruit various proteins to alter gene expression (Josling et al. 2012). As such, they are key players in transcriptional regulation, and bromodomains have recently emerged as therapeutic targets in oncogenesis and various pathological conditions (Morgado-Pascual et al. 2019; Gilan et al. 2020). Bromodomain factor 2 (Bdf2p) is one of the two bromodomain-containing proteins found in the yeast *S. cerevisiae* that recognizes acetylated lysines on histones. Bdf2p was found to establish heterochromatin boundaries and regulates the

yeast salt stress response, although the specific mechanisms that govern these processes remain elusive (Fu et al. 2013). Bdf2p is not essential for growth in *S. cerevisiae*. However, strains lacking both Bdf2p and Bdf1p are inviable (Matangkasombut et al. 2000), suggesting a partial redundancy between the functions of the two proteins. Indeed, *BDF2* overexpression can rescue the salt sensitivity and mitochondrial dysfunction of *bdf1Δ* mutants (Fu et al. 2013). This functional redundancy is further confirmed by the observation that Bdf2p can occupy sites normally bound by Bdf1p in mutants lacking Bdf1p (Durant and Pugh 2007). The absence of Bdf1p also increases the basal expression of *BDF2* (Fu et al. 2013; Volanakis et al. 2013). Interestingly, Bdf2p has been found to interact with the general Pol.II transcription factor TFIID, implicating a possible broader role in regulating transcription (Matangkasombut et al. 2000; Fu et al. 2013). These observations show that a precise regulation of *BDF2* is necessary to balance its expression relative to that of *BDF1* and during stress.

**Corresponding author:** [guillom@chem.ucla.edu](mailto:guillom@chem.ucla.edu)

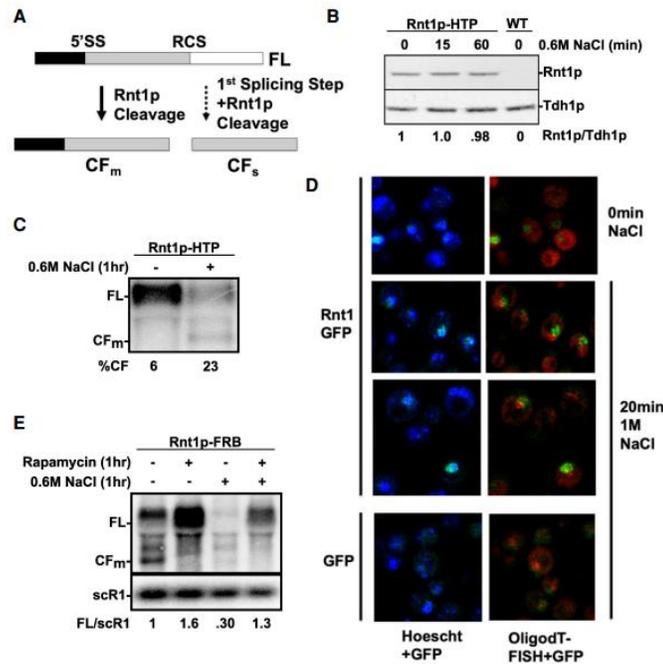
Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.078880.121>. Freely available online through the RNA Open Access option.

© 2021 Wang et al. This article, published in *RNA*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Wang et al.

*BDF2* expression has been shown to be extensively regulated post-transcriptionally in the nucleus through two distinct pathways: spliceosome-mediated decay (SMD) (Volanakis et al. 2013) and RNase III-mediated decay

(RMD) (Roy and Chanfreau 2014). During SMD, the *BDF2* mRNA undergoes the first step of splicing at a 5' splice site sequence (5'SS, Fig. 1A), but the subsequent intermediates are released and degraded, instead of proceeding



**FIGURE 1.** Rnt1p protein levels and localization remain unchanged during after salt stress. (A) Schematic representation of the cleavage of the *BDF2* mRNA by Rnt1p and by the spliceosome. Shown are the two 5'-cleavage products generated by Rnt1p-mediated decay (RMD; CF<sub>m</sub>) or by spliceosome mediated decay and RMD (CF<sub>s</sub>). These cleavage products are typically degraded by nuclear exonucleases but can be detected by inactivation of Rrp6p. The 3' fragments are not represented. (B) Western blot analysis of Rnt1p harboring an HTP carboxy-terminal tag (Granneman et al. 2009). Cells were treated with or without 0.6 M NaCl for the indicated times. Tdh1p was used as a loading control. An untagged wild-type strain was used as a negative control. Numbers below the blot show the average of three independent biological replicates. (C) Northern blot analysis of *BDF2* in the Rnt1-HTP tagged strain. Shown is a northern blot of *BDF2* using a probe spanning the *BDF2* initiation codon until the Rnt1p cleavage site (RCS; Fig. 1A) of RNAs extracted from a Rnt1-HTP tagged strain grown in normal medium or after a shift to NaCl (0.6 M, 1 h). The ncRNA scR1 was used as loading control. Numbers indicate the % of the *BDF2* cleavage fragment CF<sub>m</sub> relative to all *BDF2* transcripts. Numbers are the average of two independent experiments. (D) Localization of a Rnt1p-GFP tagged version before and after a 20 min shift to 1 M NaCl. Shown is the GFP signal from a strain expressing Rnt1-GFP, or GFP alone (bottom panel) in the indicated salt conditions, overlapped with the blue signal (DNA; Hoechst) or the red signal from the Cy3-labeled oligodT probe. (E) Northern blot analysis of *BDF2* in a strain in which Rnt1p is exported to the nucleus using the anchor away technique. Shown is a northern blot of *BDF2* showing the full length (FL) and cleavage fragment (CF) in a strain expressing Rnt1p tagged with a FKBP12-rapamycin binding domain (FRB) (Haruki et al. 2008). Strains were first treated with Rapamycin or control medium for 1 h, and then shifted to 0.6 M NaCl containing medium for 1 h or maintained in a similar medium for the control samples. Numbers indicate the levels of the full-length *BDF2* relative to the scR1 control (average of three independent replicates).

through the second splicing step (Volanakis et al. 2013). RMD is initiated by cleavage of the *BDF2* mRNA by Rnt1p, the sole representative of the RNase III family of double-stranded RNA endonucleases in *S.cerevisiae* (Elela et al. 1996; Roy and Chanfreau 2012). After cleavage in a stem-loop structure of *BDF2* by Rnt1p (the Rnt1 cleavage site or RCS, Fig. 1A), the cleavage fragments are subsequently degraded by nuclear exonucleases (Roy and Chanfreau 2014). The major upstream cleavage product (CF<sub>m</sub>, Fig. 1A) can be detected upon partial inactivation of the nuclear exosome through a deletion of its nuclear component Rrp6p (Roy and Chanfreau 2014). RMD and SMD are not completely independent, as the 5'-exon released after SMD can undergo RMD (Roy and Chanfreau 2014), resulting in a shorter cleavage fragment shown as CF<sub>s</sub> on Figure 1A. Interestingly, the two major mechanisms of *BDF2* nuclear decay are activated by different environmental stresses. During osmotic stress, RMD predominates over SMD while the opposite is true during DNA replication stress (Roy and Chanfreau 2014). The increase in the activity, or hyperactivation of RMD during salt stress results in a drastic decrease of the available pool of *BDF2* transcripts and is also responsible for the extreme salt sensitivity of *bdf1Δ* mutants (Roy and Chanfreau 2014). However, the mechanism by which RMD is hyperactivated during stress remains unclear. In this study, we provide evidence that the increased cleavage of the *BDF2* transcripts during specific stress conditions is primarily due to increased retention of the *BDF2* mRNAs within the nucleus. These results show that RMD can act as an additional layer in regulating gene expression, where transcripts are retained within the nucleus and subsequently degraded by RMD.

## RESULTS

### Rnt1p protein levels remain stable in salt stress

Previous studies have shown that the *BDF2* mRNA can be regulated by both spliceosome mediated decay (SMD) (Volanakis et al. 2013) as well as RNase III-mediated decay (RMD) (Roy and Chanfreau 2014). In the presence of high concentrations of NaCl (0.6–1 M), RMD is hyperactivated and promotes degradation of the *BDF2* transcript (Roy and Chanfreau 2014). After treatment of cells with 0.6 to 1 M NaCl, the 5'-cleavage fragment generated by RMD becomes readily detectable by northern blot analysis, while the full-length *BDF2* mRNA progressively disappears (Roy and Chanfreau 2014). To further investigate the mechanism responsible for RMD hyperactivation on the *BDF2* transcript after exposure to high concentrations of NaCl, we first analyzed Rnt1p protein levels by western blot, as increased Rnt1p expression might be responsible for increased cleavage of *BDF2*. Western blot analysis of an HTP (His6-TEV-Protein A; Granneman et al. 2009)-tagged version of Rnt1p showed no significant changes in Rnt1p protein levels (paired t-tests *P*-values >0.1; quantifications, averages, standard deviations and statistical tests applied for all the quantitative data included in the article are shown in Supplemental Table S1) or electrophoretic mobility between samples grown in standard medium versus those treated with high salt (Fig. 1B). We verified that this HTP-tagged version of Rnt1p was functional for promoting *BDF2* RMD (Fig. 1C). This suggests that the hyperactivation of RMD of the *BDF2* transcript is not a consequence of an overall increase in Rnt1p protein levels.

### Rnt1p remains localized in the nucleus during salt stress and its nuclear localization is necessary for *BDF2* RMD

Rnt1p is primarily localized in the nucleoplasm and nucleolus in normal growth conditions (Catala et al. 2004; Henras et al. 2004). However, it is unknown if the subcellular localization of Rnt1p changes during stress or under different environmental conditions. We hypothesized that if the bulk of the *BDF2* mRNA is cytoplasmic, an increase in Rnt1p cytoplasmic localization during osmotic stress may result in an increase in *BDF2* RMD. In order to visualize the localization of Rnt1p within the cell, we utilized a strain expressing a GFP-tagged version of Rnt1p, which was previously shown to be functional (Henras et al. 2004) and which can promote *BDF2* mRNA RMD during stress (see below). We did not observe a difference in the subcellular localization of Rnt1p in strains grown in normal conditions or treated with 1 M NaCl (Fig. 1D), as the major site of GFP accumulation was still detected in the nucleus, as previously shown (Henras et al. 2004). This localization was specific as it was not detected using a strain expressing a

GFP control (Fig. 1C). The primary nuclear localization of Rnt1p during salt-induced stress suggests that Rnt1p-mediated cleavage of *BDF2* transcripts during stress is likely to occur within the nucleus. Previous studies have shown that environmental stresses can cause global poly(A)<sup>+</sup> mRNA retention within the nucleus (Piper 1995; Saavedra et al. 1996; Izawa et al. 2008). Thus, it is conceivable that salt stress may cause similar global nuclear retention of poly(A)<sup>+</sup> mRNAs. In this scenario, nuclear retained transcripts would be more likely to undergo RMD. Using Cy3-labeled oligo d(T)<sub>50</sub> to visualize polyadenylated mRNAs by fluorescence in situ hybridization (FISH), a distinct pattern of nuclear poly(A)<sup>+</sup> mRNA aggregation was detected after exposure to 1 M NaCl salt stress, contrasting to the diffuse cytoplasmic pattern detected before stress (Fig. 1D). We tried to specifically detect the *BDF2* mRNAs by FISH in these conditions, but we were unable to obtain consistent data using oligonucleotide probes complementary to the *BDF2* sequence (not shown). Taken together, our data show that high salt stress does not induce an overall change in the nuclear localization of Rnt1p but results in an overall change in the localization of poly(A)<sup>+</sup> mRNAs that may contribute to promoting RMD in the nucleus.

To further show that the nuclear localization of Rnt1p is necessary for *BDF2* RMD, we depleted Rnt1p from the nucleus using the anchor away technique (Haruki et al. 2008) prior to salt stress exposure. We used a strain in which Rnt1p is tagged with a *FKBP12*-rapamycin binding domain (abbreviated as FRB), which can be used to promote the rapid export of nuclear proteins to the cytoplasm upon addition of rapamycin (Haruki et al. 2008). Anchoring away Rnt1p to the cytoplasm prior to 1 h of NaCl-induced stress was sufficient to prevent *BDF2* RMD cleavage (Fig. 1E; lane 4 vs. lane 3 paired t-test *P*-value = 0.017), further supporting the hypothesis that RMD predominantly occurs within the nucleus. Altogether, these data indicate that Rnt1p localization remains unaltered in salt stress and that RMD is dependent on the nuclear localization of Rnt1p.

### *BDF2* RMD hyperactivation can be detected in a variety of stress conditions that are known to result in mRNA nuclear retention

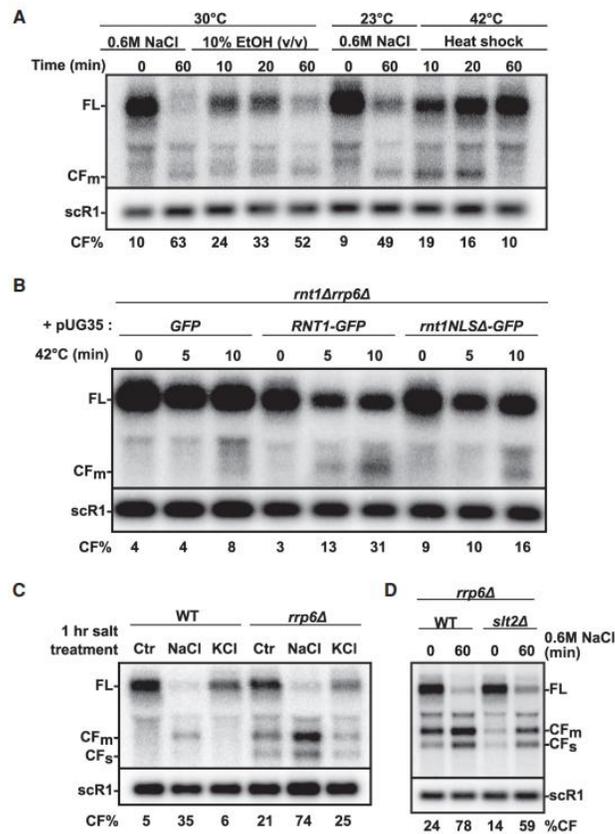
The previous data showed that nuclear localization of Rnt1p is necessary for the cleavage of *BDF2* transcripts, suggesting that an increase of *BDF2* mRNA nuclear retention during salt stress may be a primary mechanism for RMD hyperactivation. To further test this hypothesis, we subjected *S. cerevisiae* to various stress conditions and analyzed *BDF2* transcripts by northern blotting. RMD hyperactivation during these stress conditions was visualized either by estimating the amount of full-length mRNA remaining or by calculating the overall % of cleavage fragment (CF) over the overall *BDF2* transcripts population [%CF = CF/(CF + FL) × 100]. Ethanol or heat shock stress are known

Wang et al.

to cause selective retention of bulk mRNAs within the nucleus and the rapid export of stress responsive transcripts (Piper et al. 1994; Saavedra et al. 1997; Izawa et al. 2008). Strikingly, we detected RMD hyperactivation of *BDF2* in cells treated with either ethanol or heat shock (Fig. 2A) but the kinetics and final effect on full-length *BDF2* levels varied. Ethanol treatment resulted in a global decrease of the *BDF2* mRNA and an accumulation of the main cleavage fragment CFm after 1 h. However, during heat shock the cleavage fragment was detected at its peak within the first 20 min of treatment and the level of the *BDF2* FL mRNA was not strongly affected. No cleavage fragment was detected after an hour of heat shock, as the cells may have recovered from the stress. This is consistent with RNA-seq analysis done after 45 min of heat shock (Wang et al. 2020), which showed that *BDF2* full-length levels were similarly to control samples after prolonged heat-shock. Interestingly, a complete loss of the full length *BDF2* transcript was not detected during salt treatment at 23°C, as compared to salt treatment at 30°C (compare FL between lanes 7 and 2, Fig. 2A). This suggests that the steady state growth temperature may influence the cleavage activity of Rnt1p as well. Overall, these data using heat shock and ethanol stress treatments suggest that the RMD hyperactivation seen on *BDF2* mRNAs may be due to its nuclear retention, which has been shown to occur globally on mRNAs in these stress conditions (Piper et al. 1994; Saavedra et al. 1997; Izawa et al. 2008).

To test the hypothesis that *BDF2* RMD during heat-shock requires Rnt1p nuclear localization, we analyzed *BDF2* during heat shock in a *rnt1Δ rrp6Δ* double deletion mutant strain expressing either GFP alone, the *RNT1*-GFP fusion used for the localization studies shown above, or a *RNT1*-GFP carrying a deletion of the nuclear localization signal of Rnt1p (NLSΔ), which results in a large fraction of Rnt1p delocalized to the cytoplasm (Henras et al. 2004). Cells expressing the Rnt1-GFP fusion exhibited a decrease

of the FL form of *BDF2* after heat-shock, with an increased accumulation of the *BDF2* cleavage fragment compared to cells expressing GFP alone (Fig. 2B). This result shows that the Rnt1-GFP fusion is functional for RMD during heat



**FIGURE 2.** Analysis of *BDF2* RMD in various stress conditions. (A) Northern blot analysis of *BDF2* following exposure to ethanol (10%), heat shock (42°C), or salt stress conditions (0.6 M NaCl). scR1 was used as a loading control. The migration of the full-length (FL) and of the main cleavage fragment (CFm) is indicated. Numbers indicate the % of the *BDF2* cleavage fragments relative to all *BDF2* transcripts. The numbers shown are the average of three independent biological replicates, except for lane 5 (60 min ethanol treatment) for which only two independent replicates were obtained. (B) Northern blot analysis of *BDF2* in an *rrp6Δ rnt1Δ* strain expressing GFP alone, Rnt1-GFP or Rnt1-GFP partially delocalized to the cytoplasm (NLSΔ) from pUG35 plasmids (Henras et al. 2004) in normal and heat-shock conditions. Numbers indicate the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts calculated from the northern blot shown on the picture which is the only biological replicate for this experiment. (C) Northern blot analysis of *BDF2* in control medium or following a 1 h exposure to 0.6 M NaCl or 0.6 M KCl. Numbers indicate the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts (average of three independent replicates). (D) Northern blot analysis of *BDF2* during NaCl-induced stress in the *rrp6Δ* and *rrp6Δ slt2Δ* mutants. Legends as in Figure 2A. Numbers indicate the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts based on the average of three independent biological replicates.

shock. In contrast, cells expressing the Rnt1p–NLSΔ mutant showed a delay in the accumulation of the *BDF2* cleavage fragment, with none detectable after 5 min of heat shock treatment, and a reduced accumulation after 10 min compared to the wild-type Rnt1–GFP (Fig. 2B). RMD was not detected in the strain expressing GFP alone, showing that the band detected upon heat-shock is the result of Rnt1p cleavage. Even though this experiment was only performed once, this result shows that a fully functional Rnt1p localization signal is required for optimal RMD of *BDF2* during heat shock. These experiments also show that, as opposed to salt stress, heat shock results in a transient activation of RMD.

We further explored the type of ionic stress that can mediate RMD of *BDF2*. We found that the addition of high KCl concentrations (0.6 M) did not activate RMD, indicating that RMD hyperactivation is specific to the type of ionic stress (Fig. 2C). A similar result was obtained when nuclear exosome activity was impaired by the deletion of Rrp6p to increase detection of the CF (Fig. 2C). Altogether, these results demonstrate that *BDF2* RMD can be hyperactivated in a variety of stress conditions that are known to specifically induce mRNA nuclear retention.

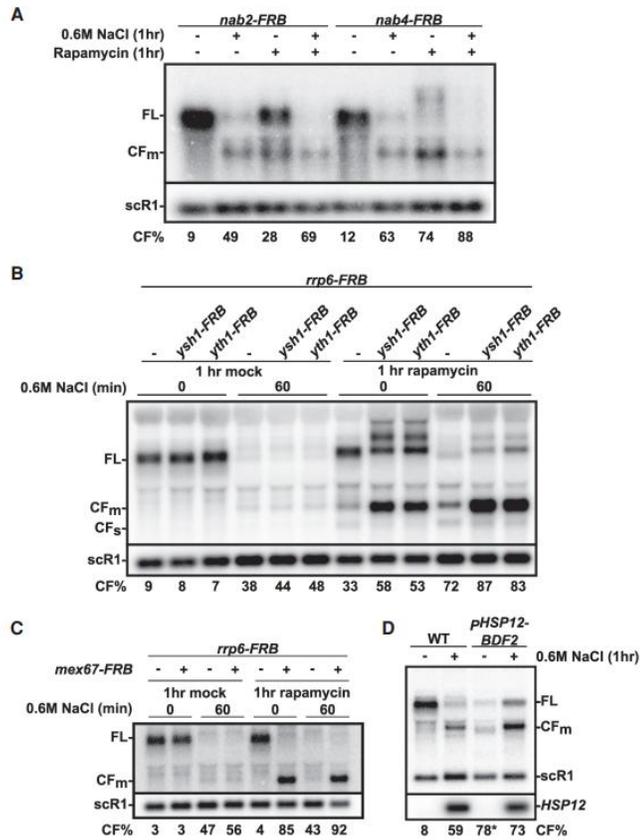
***BDF2* RMD hyperactivation is reduced in the *slt2Δ* mutant but can be uncoupled from extracellular stress by anchoring away mRNA biogenesis or export factors**

Specific stress conditions, including salt stress and heat stress can promote the nuclear retention of non-heat shock mRNAs as a mechanism to rapidly change gene expression (Carmody et al. 2010). Slt2p (also referred to as Mpk1p) is a key protein kinase involved in the stress response (Kim and Levin 2011), which is required for the retention of non-heat shock mRNAs during heat shock conditions (Carmody et al. 2010). We used a strain carrying a deletion of the *SLT2* gene to determine if stress-induced hyperactivation of RMD is caused by *BDF2* nuclear retention through the Slt2p pathway. We performed this experiment in an *rrp6Δ* background to maximize the detection of the *BDF2* cleavage fragment. The absence of Slt2p significantly reduced *BDF2* RMD activation during NaCl-induced stress (lane 4 vs. lane 2; paired t-test *P*-value = 0.008; Supplemental Table S1), as shown by the increased accumulation of the full-length *BDF2* mRNA and the relative decrease of the cleavage product during stress compared to the wild-type control (Fig. 2D). However, the absence of Slt2p did not fully inhibit RMD activity on *BDF2* during salt stress. This result suggests that Slt2p might contribute to the increased retention of *BDF2* during stress, but that the *BDF2* mRNA could also be retained through a different pathway not defined by Slt2p. For example, specific stress signals may impact the mRNP formation, processing and export of several transcripts differently (Izawa et al. 2008).

To further test the hypothesis that the hyperactivation of RMD on *BDF2* is primarily due to its nuclear retention during stress, we tested the effect of globally inhibiting mRNA export to the cytoplasm independently from any extracellular stress. We used the anchor-away technique (Haruki et al. 2008) to rapidly deplete from the nucleus several proteins that are involved directly or indirectly in the export of mRNAs to the cytoplasm. We first focused on the poly(A) binding protein Nab2p and on the 3'-end processing factor Nab4p (Hrp1p), as defects in poly(A) tail binding or formation have been shown to inhibit mRNA export (Hammell et al. 2002). As a control, salt treatment of strains in which Nab2p and Nab4p were FRB-tagged (but without adding rapamycin which promotes the anchor away process) resulted in a reduction of the full length *BDF2* transcripts and in the appearance of the cleavage product CFm (Fig. 3A). Strikingly, cleavage of *BDF2* transcripts was detected in the absence of any salt stress after anchoring away Nab2p or Nab4p (Fig. 3A). The size of the fragment detected upon anchoring away Nab2p or Nab4p in the absence of stress was identical to the size of the fragment detected during salt stress (Fig. 3A), which showed that RMD was activated upon anchoring away Nab2p or Nab4p. The extent of *BDF2* cleavage differed, with Nab4p resulting in a stronger accumulation of the CFm and in a very strong reduction in the levels of the FL form of *BDF2*; however, this reduction was not solely due to RMD, as long extended forms could also be detected due to transcription termination defects arising from defective 3'-end formation when Nab4p is inactivated (Minvielle-Sebastia et al. 1998). Nevertheless, these results suggested that the inhibition of mRNA export triggered by defects in poly(A) formation or binding is sufficient to recapitulate the salt stress-induced RMD hyperactivation of *BDF2*. However, anchoring away Nab2p or Nab4p did not result in an activation of RMD to the same extent as that detected during salt stress, and a further loss of the full-length transcript was observed when cells were exposed to salt stress prior to Rapamycin treatment (Fig. 3A). This additive effect may either indicate a synergy between the retention of unprocessed RNAs and the retention of mature RNAs during stress for RMD activation, or that the anchoring away of either of these two factors may not completely abolish mRNA export.

To further explore the hypothesis that inhibiting 3'-end processing can result in *BDF2* RMD hyperactivation, we constructed strains in which the core cleavage and polyadenylation (CPA) factors Yth1p and Ysh1p can be anchored away from the nucleus, concomitantly with Rrp6p (Fig. 3B). This double anchor away strategy was used to allow a better detection of the *BDF2* cleavage product as anchoring away Rrp6p strongly stabilizes nuclear exosome targets (Roy et al. 2016), possibly because other subunits of the exosome might be anchored away concomitantly. Strikingly, anchoring away both Rrp6p and Ysh1p or Yth1p resulted

Wang et al.



**FIGURE 3.** Anchoring away mRNA biogenesis or mRNA export factors triggers RMD hyperactivation of *BDF2* in the absence of stress. (A) Northern blot analysis of *BDF2* in strains expressing FRB-tagged versions of Nab2p, and Nab4p in salt stress conditions and/or in conditions triggering nuclear export of these factors upon addition of Rapamycin. Cells grown to log phase were split in half, with one kept in normal medium and the other half treated with 0.6 M salt. After an hour of salt treatment, each culture was split in half again and treated with either 1  $\mu$ g/mL of rapamycin or its vehicle solvent (90% ethanol and 10% Tween-20) for an additional hour. scR1 was used as a loading control. The *BDF2* full-length (FL) and cleavage fragments (CF<sub>m</sub>) are shown, with the scR1 loading control. Numbers indicating the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts are the average of three independent biological replicates. (B) Northern blot analysis of *BDF2* in strains expressing FRB-tagged versions of Rrp6p and of the cleavage and polyadenylation factors Yth1p or Ysh1p, in salt stress conditions and/or upon addition of Rapamycin. Legends are as in Figure 3A. Numbers indicating the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts are the average of three independent biological replicates, except for lanes 1, 4, 7, and 10 (two independent replicates), and 2 and 3 (one replicate). (C) Northern blot analysis of *BDF2* in a strain expressing FRB-tagged versions of Rrp6p and of the mRNA export factor Mex67p in salt stress and/or upon addition of Rapamycin. Legends are as in Figure 3A. Numbers indicating the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts are the average of three independent biological replicates. (D) Northern blot analysis of the *BDF2* and *HSP12* mRNAs in a wild-type strain and in a strain expressing *BDF2* from the *HSP12* promoter in normal and in salt stress conditions. Legends as in Figure 3A. Numbers indicating the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts are the average of ten (lanes 1 and 2) or three (lanes 3 and 4) independent biological replicates.

in a strong accumulation of the *BDF2* cleavage product (Fig. 3B) in the absence of any extracellular stress. The fragment detected upon anchoring away Ysh1p or Yth1p and Rrp6p in the absence of stress comigrated with the fragment detected during salt stress treatment alone (Fig. 3A), which showed that this fragment was generated by Rnt1p cleavage. The accumulation of the *BDF2* cleavage fragment and the decrease of the full-length mRNA were exacerbated when strains were exposed to salt stress prior to anchoring away these CPA factors. Similar to what was observed with Nab4p, anchoring away Ysh1p and Yth1p also resulted in the appearance of extended species of *BDF2*, likely because of the transcription termination defects resulting from CPA inactivation. These results show that inactivation of different CPA factors can promote the hyperactivation of *BDF2* RMD.

To formally demonstrate that inhibition of mRNA nuclear export can trigger RMD hyperactivation, we performed nuclear depletion of the essential mRNA export factor, Mex67p concomitantly with Rrp6p. Mex67p is a core member of the mRNA export complex (Hurt et al. 2000) and its nuclear depletion through the anchor-away technique completely abolishes poly(A)<sup>+</sup> mRNA export (Haruki et al. 2008). Strikingly, co-nuclear depletion of Mex67p and Rrp6p fully reproduced *BDF2* RMD hyperactivation (Fig. 3C) in the absence of extracellular stress. Nuclear depletion of Mex67p resulted in a complete loss of the full length *BDF2* transcripts, a phenotype stronger than the nuclear depletion of Nab2p or of the 3'-end processing factors Nab4p, Yth1p, or Ysh1p. The complete cleavage of the *BDF2* transcripts was further confirmed by the observation that no additional cleavage product accumulated when the anchored-away strains were treated with salt stress (Fig. 3C). Overall, these results indicate that the subcellular localization of *BDF2* mRNAs plays a pivotal role in determining their

degradation fate through RMD, and that inhibition of mRNA nuclear export by triggering export of key mRNA biogenesis or export factors can fully recapitulate RMD hyperactivation of *BDF2* in the absence of extracellular stress.

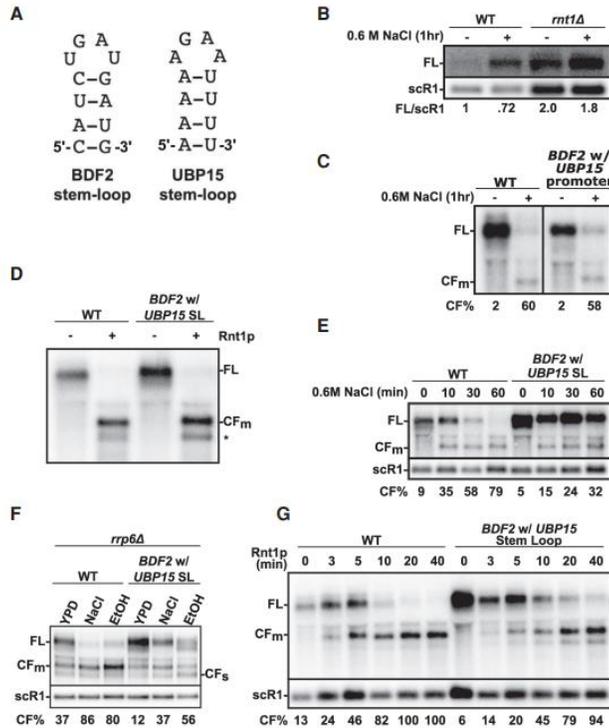
#### Promoter identity is not required to promote BDF2 RMD during stress

Elements within gene promoters can affect the stability of mRNAs expressed from these promoters (Trcek et al. 2011; Catala and Abou Elela 2019). Promoter-dependent regulation of mRNA stability is controlled through the cotranscriptional recruitment of several proteins that initiate mRNA decay. Rnt1p itself can be recruited to the promoter before recognizing the Rnt1p cleavage site (RCS) within the ORF (Catala and Abou Elela 2019). To determine if *BDF2* RMD relies on the recruitment of Rnt1p to the *BDF2* promoter, we swapped the endogenous *BDF2* promoter with that of *HSP12*. The *HSP12* mRNA does not undergo RMD, and we hypothesized that this promoter would be unable to recruit Rnt1p despite containing many stress response elements (STRE) (Varela et al. 1995). Furthermore, the *HSP12* promoter is strongly activated under salt stress, allowing us to investigate the RMD-mediated regulation of the *BDF2* transcript generated from this promoter ( $\rho$ *HSP12*-*BDF2*). The *BDF2* mRNA generated from  $\rho$ *HSP12*- was still targeted by Rnt1p for rapid cleavage under salt stress (Fig. 3D). In fact, the strong induction of the *HSP12* promoter under salt stress generated more transcripts to be targeted by RMD, resulting in an increase of accumulation of the cleavage fragment compared to wild type post-salt stress. RMD was however not sufficient to completely eliminate the highly expressed  $\rho$ *HSP12*-*BDF2* transcript, possibly due to the known ability of STRE to promote rapid export in heat shock through recruitment of the mRNA export factor Mex67p by Hsf1p (Zander et al. 2016). We conclude that the RMD hyperactivation on the *BDF2* transcript does not require recruitment of Rnt1p to the endogenous *BDF2* promoter.

#### The identity of the stem-loop cleaved by Rnt1p influences RMD efficiency during stress

Previous studies have demonstrated that specific structural features of the stem-loop recognized by Rnt1p may affect cleavage efficiency (Comeau et al. 2016). To further investigate how the stem-loop present in *BDF2* may affect RMD hyperactivation, we replaced the endogenous *BDF2* RCS with that of *UBP15* (Fig. 4A). The *UBP15* mRNA does not undergo RMD hyperactivation in salt stress (Fig. 4B) and replacing the *BDF2* promoter by the *UBP15* promoter does not abrogate RMD (Fig. 4C), which showed that the lack of cleavage of *UBP15* in vivo during stress is not due to its endogenous promoter, which could have promoted its ex-

port from the nucleus as shown for other stress induced genes (Zander et al. 2016). However, the *BDF2* mRNA containing the stem-loop (SL) present in *UBP15* was fully cleaved when incubated for 1 h with recombinant Rnt1p in vitro (Fig. 4D). Incubation of total RNAs containing the *BDF2* hybrid transcripts containing the *UBP15* SL with recombinant Rnt1p resulted in detection of a cleavage product with a migration identical to that detected upon in vitro cleavage of the wild-type *BDF2* (Fig. 4D). This indicates that the hybrid *BDF2* transcripts can be indeed targeted and cleaved by Rnt1p in the same stem-loop structure. We note that the *UBP15* terminal loop also induced the cleavage at an additional, minor site (asterisk on Fig. 4D). Based on these in vitro cleavage data, it is unclear whether the absence of RMD activity on *UBP15* in vivo is due to the identity of its RCS or to other factors that might prohibit cleavage in vivo. To further investigate the impact of the RCS identity on RMD, we analyzed the levels of wild-type *BDF2* and of the mutant containing the *UBP15* SL during stress. The *BDF2* hybrid transcript containing the *UBP15* SL was less efficiently down-regulated by RMD during a kinetic of salt exposure (Fig. 4E), with significant decreases of RMD efficiency compared to the WT based on the %CF (all time points paired t-tests P-values <0.05; Supplemental Table S1). To better characterize the impact of the *UBP15* stem-loop on *BDF2*, we introduced this mutation in the context of the *rrp6Δ* mutant, and we analyzed *BDF2* levels in control medium (YPD) or after 1 h of 0.6 M NaCl or 10% ethanol treatment (Fig. 4F). In rich medium, the presence of the *UBP15* SL resulted in a substantial increase of the FL form and a decrease of both CFm and CFs (Fig. 4F), reflecting a significant decrease in Rnt1p cleavage activity in normal growth conditions (lane 1 vs. 4; t-test P-value = 0.005; Supplemental Table S1). The *UBP15* stem-loop also decreased the hyperactivation of *BDF2* RMD during stress, based on the %CF ratios indicated and as shown by a substantial amount of the full length *BDF2* hybrid transcript remaining after an hour of salt or ethanol treatment compared to the wild-type *BDF2* (paired t-test P-values: NaCl treatment = 0.001; ethanol treatment = 0.05). To further characterize the impact of the *UBP15* stem-loop on Rnt1p cleavage activity, a kinetic analysis of in vitro cleavage was performed using purified Rnt1p and total RNAs as a source of substrate RNA. As shown in Figure 4G the rate of in vitro cleavage of the hybrid transcript was slower than that observed for the wild-type *BDF2* transcript (Fig. 4G). Although this time course experiment was performed only once, the decreased cleavage efficiency is consistent with the decreased cleavage activity detected in vivo in Figure 4E,F. Inefficient cleavage of this substrate compared to the WT could be due to the fact that the *UBP15* stem-loop is rich in A-U base pairs near the terminal tetraloop, which is unusual for Rnt1p substrates (Chanfreau 2003). Taken together, our data demonstrate that specific features of the RCS can affect RMD



**FIGURE 4.** The identity of the Rnt1p cleavage site (RCS) influences RMD hyperactivation in stress conditions. (A) Predicted secondary structures of the terminal stem-loops present at the top of the RCS of *BDF2* and *UBP15* using Mfold (Zuker 2003). (B) Analysis of *UBP15* mRNA expression in WT and *mtf1Δ* cells in normal growth conditions and upon salt stress treatment. Shown is a northern blot analysis of *UBP15* in the corresponding strains. scR1 was used as a loading control. This experiment was performed once and the ratios indicate the level of full-length *UBP15* relative to scR1 based on the northern blot shown. (C) Northern blot analysis of *BDF2* in normal or NaCl-stress conditions for the endogenous *BDF2* gene or the gene expressing *BDF2* from the *UBP15* promoter. The FL and main cleavage fragments are indicated. This experiment was performed once and the %CF was calculated from the northern blot shown on this figure. (D) In vitro cleavage assay of *BDF2* transcripts containing the RCS of *UBP15*. Shown is a northern blot of total RNAs from the corresponding strains after incubation in buffer or with recombinant Rnt1p for 1 h. The location of the main cleavage fragment (CF<sub>m</sub>) is indicated. The asterisk indicates a secondary, minor cleavage event detected for the *BDF2*-*UBP15* SL substrate. (E) Northern blot of wild-type *BDF2* and *BDF2* transcripts harboring the RCS from *UBP15* after exposure to 0.6 M NaCl stress. Legends as in Figure 2. Numbers indicating the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts are the average of three independent biological replicates. (F) Northern blot of *BDF2* in an *rrp6Δ* mutant expressing the wild-type *BDF2* or a *BDF2* mutant with the *UBP15* stem-loop in normal medium, or after exposure to 0.6 M NaCl or 10% ethanol for 1 h. Numbers indicating the % of the *BDF2* cleavage fragment relative to all *BDF2* transcripts are the average of three independent biological replicates. (G) Time course of in vitro cleavage of the wild-type *BDF2* and *BDF2* RNA with the *UBP15* terminal stem-loop by recombinant Rnt1p. Total RNAs extracted from the wild-type strain or from a strain expressing *BDF2* with the *UBP15* terminal stem-loop were incubated with recombinant Rnt1p for the indicated times. The reactions were stopped, the RNAs repurified and analyzed by northern blot. The numbers indicate the % of *BDF2* cleavage [(CF/CF + FL) × 100] relative to time zero for each substrate for the time course experiment shown on this figure.

activity during normal and stress conditions and that optimal degradation of *BDF2* during stress requires both its specific stem-loop structure and inhibition of RNA export, but is not dependent on expression of *BDF2* from its natural promoter.

## DISCUSSION

Bdf2p is a bromodomain protein involved in multiple aspects of epigenetic control through its ability to recognize acetylated histones. We previously showed that the *BDF2* transcript is subject to both RMD and SMD pathways which limit and regulate its expression (Roy and Chanfreau 2014). This regulation is of particular importance during stress conditions, where a rewiring of transcription and translation is necessary to promote cellular fitness during stress. In high salt stress, an increase in activity of RMD on *BDF2* causes its transcript to become completely undetectable. Repression of *BDF2* in stress conditions is necessary for the robust expression of the stress responsive gene *GPH1*, and potentially other stress responsive genes as well (Roy and Chanfreau 2014). Despite the significance of regulating the expression of *BDF2*, the mechanism that triggers RMD hyperactivation during salt stress was not clearly understood. In this study, we demonstrate that the increase of RMD activity on *BDF2* during stress does not arise from a direct change in the expression of Rnt1p or its localization within the cell. Rather, our study suggests that the nuclear retention of *BDF2* transcripts during stress causes RMD hyperactivation. A global block of nuclear export of mRNAs induced by stress conditions such as heat or ethanol shock, or by nuclear depletion of specific mRNA biogenesis or export factors, can reproduce salt stress-induced RMD hyperactivation on *BDF2*. We also show that inactivation of the key stress response factor Slt2p reduces the efficiency of *BDF2* RMD during stress. We interpreted this reduction in RMD in

the *slt2Δ* mutant as a consequence of the role of Slp2p in mediating stress-induced mRNA nuclear retention (Carmody et al. 2010). However, we do not know if the reduction of RMD efficiency detected in this mutant for RMD is dependent on its kinase and signaling activity, as Slp2p has kinase-independent roles during stress (Kim and Levin 2011). Taken together, these results show that the primary mechanism underlying the hyperactivation of RMD for *BDF2* is the inhibition of mRNA export and the retention of transcripts in the nucleus. RMD hyperactivation is independent of the identity of the *BDF2* transcript promoter region, as switching its promoter to that of *HSP12* or *UBP15* does not prevent RMD. This suggests that promoter-dependent recruitment of Rnt1p, which is required for RMD of other transcripts (Catala and Abou Elela 2019) does not play a role in the hyperactivation of RMD of *BDF2* observed in stress conditions.

We further provide evidence that the identity of the Rnt1p cleavage stem-loop (RCS) within *BDF2* further influences its susceptibility to cleavage, as replacing the endogenous *BDF2* terminal stem-loop by that of *UBP15* decreases *BDF2* RMD in vivo, and Rnt1p cleavage efficiency in vitro. This supports previous evidence that the nucleotide base-pairing of the product termini can determine the Rnt1p substrate reactivity (Comeau et al. 2016). The effect of the *UBP15* stem-loop on Rnt1p cleavage might be linked to its stretch of A-U base pairs near the terminal tetraloop (Fig. 4A), which is not usually found in natural Rnt1p substrates (Chanfreau 2003). It is unclear whether certain stress conditions, such as heat shock or salt stress, may alter the structure of the RCS in vivo and influence its cleavage efficiency by Rnt1p. High temperatures or specific salt conditions may contribute to the stabilization or destabilization of the *BDF2* RCS and thereby modulate cleavage efficiency. This may contribute to explain some of the differences in *BDF2* RMD efficiency detected between heat shock treatment, ethanol or salt stress. Alternatively, it is possible that these different stress conditions affect mRNA export out of the nucleus differently, or that they could impact degradation pathways which facilitate the detection of the *BDF2* cleavage products.

Altogether, our results indicate that Rnt1p may help regulate expression of specific genes through the cleavage and degradation of specific substrates based on their localization and stem-loop structures. The nuclear export of many mRNAs is inhibited during stress, which provides the opportunity for these mRNAs to be targeted by Rnt1p for degradation in these conditions to differing extents, depending on substrate reactivity. Previous studies have shown that blocking the nuclear export of mRNAs by nuclear depletion of Mex67p results in the rapid degradation of newly synthesized RNAs (Tudek et al. 2018). A similar effect was observed for polyadenylated RNAs when the poly(A) binding protein Nab2p was depleted from the nucleus (Schmid et al. 2015). Strikingly, anchoring

away both of these factors is sufficient to promote *BDF2* RMD in the absence of stress (Fig. 3). Therefore, it is possible that Rnt1p may play a role in the general degradation mechanism previously described by the Jensen group upon nuclear depletion of these factors, and that Rnt1p-mediated cleavage may contribute to the elimination of specific mRNAs containing compatible cleavage sites during stress-induced nuclear retention of mRNAs. However, not all mRNAs sequestered in the nucleus during stress or after nuclear depletion of CPA or export factors contain stem-loop structures compatible with Rnt1p cleavage, and it is likely that the sequence content of mRNAs has evolved to either promote RMD of mRNAs that are not required or perhaps detrimental during stress conditions.

## MATERIALS AND METHODS

### Yeast strains and plasmids

All strains used in this study were derived from BMA64, HHY168 or as described below:

- Strains previously described or from external sources:
  - BMA64 (Chanfreau et al. 1998): MATa; *ura3-1*; *trp1Δ2*; *leu2-3,112*; *his3-11,15*; *ade2-1*; *can1-100*
  - mt1Δ*: BMA64, *mt1Δ::TRP* (Chanfreau et al. 1998)
  - BY4742: MATα *his3Δ1* *leu2Δ0* *lys2Δ0* *ura3Δ0* (Open Biosystems)
  - HHY168 (Haruki et al. 2008) MATα *tor1-1* *fpr1::NAT* *RPL13A-2×FKBP1 2::TRP1* *ade 2-1* *trp1-1* *can1-100* *leu2-3,112* *his3-11,15* *ura3* *GAL* *psi+* (Haruki et al. 2008)
  - yCW2: MATa *his3Δ1* *leu2Δ0* *met15Δ0* *ura3Δ0* *RNT1-HTTP::KIURA3* (gift of D. Tollervy).
  - yCW6: BY4742, *rrp6Δ::hphMX6* (Wang et al. 2020)
  - yCW17: yCW6, *slt2Δ::kanMX6* (Wang et al. 2020)
- Strains generated in this study:
  - yCW4: MATα *tor1-1* *fpr1::hphMX6* *RPL13A-2×FKBP12::TRP1* *ade 2-1* *trp1-1* *can1-100* *leu2-3,112* *his3-11,15* *ura3* *GAL* *psi+*
  - yCW5: yCW4 with *RNT1-FRB::kanMX6*
  - yCW7: yCW4 with *NAB2-FRB::kanMX6*
  - yCW8: yCW4 with *NAB4-FRB::kanMX6*
  - yCW9: yCW4 with *YTH1-FRB::kanMX6* *RRP6-FRB::his3MX6*
  - yCW10: yCW4 with *YSH1-FRB::kanMX6* *RRP6-FRB::his3MX6*
  - yCW11: yCW4 with *MEX67-FRB::kanMX6* *RRP6-FRB::his3MX6*
  - yCW12: BMA64 with *pHSP12-BDF2*
  - yCW13: BMA64 with *BDF2* with *UPB15 SL*
  - yCW14: BMA64 with *BDF2* with *UPB15 SL* *rrp6Δ::hphMX6*
  - yCW18: BMA64 *mt1Δ::TRP*, *rrp6Δ::hphMX6*
  - yCW19: BMA64 with *BDF2* with *UBP15* promoter

Strains were constructed using a high efficiency transformation method (Gietz and Schiestl 2007). Mutations within the *BDF2* transcripts were constructed through the *delitto perfetto*

Wang et al.

approach (Storici and Resnick 2006). The Rnt1p target stem-loop within the *BDF2* transcript (ChrIV:332367-80) was replaced with the CORE integration cassette, consisting of the *URA3* and *KanMX6* genes. Successful transformants were selected through their resistance to G418, and further confirmed through PCR. Afterwards, the CORE integration cassette was replaced with various Rnt1p target stem-loops using the transformation protocol as described above. Successful transformants were initially selected on the basis for their ability to grow on 5-Fluoroorotic acid (5-FOA) due to the loss of *URA3*, inability to grow on G418 due to the loss of *kanMX6*, and then further confirmed through PCR and Sanger sequencing (Laragen, Inc.).

Anchor away strains were created in a modified HHY168 (Haruki et al. 2008) background where the *natMX6* marker was replaced with the *hphMX4* marker amplified from pAG32 (Goldstein and McCusker 1999). The genes of interest were carboxy-terminally tagged with the rapamycin binding domain (FRB) using the transformation method as described. Plasmids expressing Rnt1-GFP and Rnt1-GFP with a deletion of the nuclear localization sequence are described in Henras et al. (2004).

#### Oligonucleotides used for mutagenesis (F and R stand for forward and reverse primers)

BDF2 Promoter Core oligonucleotides:  
 F:AGAGGCGAAAAAGAGTGCAACGTCACAACGCTAAAA  
 GAGAGCTCGTTTTGACACTGG  
 R:AGCCGCGAGGTTTATTTCGCTCAATCTGTTTGTTCAGT  
 TCCTTACCATTAAGTTGATC  
 HSP12::BDF2 promoter swap:  
 F:AATGGAGTGAAGCAGGCAGGGTGACCCTCTAGCTAAAA  
 AA GATCCCCTAACGGCCGAGC; R:CAGAATGTGCGTG  
 TCTTGATCCATGTTAGTACGAGACAT TGTTGATTTAGT  
 TTTTTT  
 UBP15::BDF2 promoter swap:  
 F:AGAGGCGAAAAAGAGTGCAACGTCACAACGCTAAAA  
 GA CAAGAGAGCAGTAGTAAGAG  
 R:CAGAATGTGCGTGCTTGTATCCATGTTAGTACGAGACA  
 T TGTTGTTTGAAGAGACTAA  
 BDF2 Rnt1 cleavage site Core insertion:  
 F:GGATTGAGATCCTTGAAGAGGATACTATTCTTCTTCATAT  
 GAGCTCGTTTTGACACTGG  
 R:TAGTTATATCGTTTTTATTATGTCCTTCATCATCATATTCTC  
 CTTACCATTAAGTTGATC  
 UBP15 stem-loop swap:  
 R:ATATTTTGAATGAAATTGAAACGGATTGAGATCTTGAAGA  
 GGATGCAAAATTTGGTCTCGGACAAAAGAATTTTCAAAG  
 F:TTTGTCCAAATATTGAATAGCCGGATTAGTTATATCGTTT  
 TCATTACTTAATTTGATCCTTGAAAATCTTTTTGTCCG

#### Yeast media and growth conditions

All strains were grown in YPD (1% yeast extract, 2% peptone, and 2% dextrose) or minimal media (0.67% w/v yeast nitrogen base, 2% w/v dextrose, and 0.2% w/v amino acid mixture) at 30°C unless noted otherwise. For protein and RNA extractions, 50 mL of culture were harvested at OD<sub>600</sub> 0.4–0.6 by centrifugation at 4000 rpm (Sigma Rotor 11030) for 1.5 min and transferred to 2 mL screw capped Eppendorf tubes. The cells were pelleted

and flash frozen in liquid nitrogen. For experiments involving treatment involving anchor away and salt stress conditions, specific conditions are described in the results or legends of each figure. For heat shift experiments, cells were grown to exponential phase at 23°C before equal volumes of 61°C preheated YPD were added to bring the temperature to 42°C. The cultures were immediately harvested at the indicated times. For ethanol treatment, cells were grown to exponential phase before equal volumes of YPD or YPD containing 20% ethanol (v/v) were added, and then harvested at the indicated times.

#### Yeast RNA extraction and northern blot analysis

For most experiments, northern blot analysis was performed in triplicate (unless indicated otherwise) on purified total RNAs as described previously (Roy and Chanfreau 2014). Briefly, RNAs were extracted using phenol-chloroform, precipitated with ethanol and sodium acetate, and resuspended in water. A total of 5 µg of glyoxylated RNAs were run on a 1.8% agarose gel and transferred to nylon membranes for probing.

#### Riboprobe and oligonucleotide probe synthesis for northern blotting analysis

The radiolabeled riboprobes were transcribed in vitro using T3 RNA polymerase (Promega) as described previously (Roy and Chanfreau 2014). The *BDF2* riboprobe hybridizes from the beginning of the open-reading frame until the Rnt1p cleavage site.

Templates used were synthesized by PCR using the primers: BDF2 F: GCACATTCTGCTTTACTGGCAGC and BDF2 T3R: GGCTAAATTAACCTCACTAAAGG TTCAAGATCTGAATCCGTTT.

The SCR1 Oligonucleotide used for the loading control: ATCCCCGGCCCTCCATCAC was radiolabeled using  $\gamma$ -<sup>32</sup>P-ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's protocol.

#### In vitro Rnt1p cleavage assay

Recombinant Rnt1p was purified as described previously (Henras et al. 2005). In vitro cleavage reactions were performed in 50 µL reactions consisting of 50 µg of total yeast RNA, 10 pmol of purified recombinant Rnt1p in Rnt1p cleavage buffer (30 mM Tris pH 7.5, 150 mM KCl, 5 mM spermidine, 200 mM MgCl<sub>2</sub>, 0.1 mM DTT, 0.1 mM EDTA). The reactions were incubated at 30°C and halted by the addition of 150 µL of RNA buffer at the times indicated. The reactions were then purified through using phenol-chloroform. Briefly, 200 µL of phenol: chloroform: isoamyl alcohol were added to the samples. The samples were vortexed for 1 min and spun down for 2 min at 15,000 rpm. The top aqueous layer was added to a fresh Eppendorf tube containing 1 mL ethanol, 40 µL 3 M sodium acetate pH 5.2, and 1 µL of GlycoBlue (Ambion). Precipitation of the RNA was facilitated through incubating the samples at –80°C for 30 min and then pelleted by centrifugation at 15,000 rpm for 10 min. The pellets were washed with 200 µL of 70% ethanol and resuspended in 15 µL of nuclease-free water.

### Protein extraction and western blot analysis

Western blot analysis was performed using protein extracts from an Rnt1p-HTP strain kindly provided by D. Tollervy (U. Edinburgh). The HTP carboxy-terminal tag (Granneman et al. 2009) consists of a 6-His sequence, followed by a TEV protease cleavage site and a Protein A sequence (PMID: 19482942). Total protein was prepared from the Rnt1p-HTP strain and an untagged control strain grown to mid-log phase in 50 mL YPD liquid medium and treated with 0.6 M NaCl for 0, 15, and 60 min. The culture was harvested by centrifugation for 5 min at 3500 rpm, washed with ddH<sub>2</sub>O, and resuspended in lysis buffer (200 mM Tris-HCl pH 8.0; 320 mM Ammonium sulfate; 5 mM MgCl<sub>2</sub>; 10 mM EGTA pH 8.0; 20 mM EDTA pH 8.0; 1 mM DTT; 20% glycerol; 1 mM PMSF; 2 mM benzimidazole HCl and protease inhibitor cocktail). The sample was vortexed with glass beads (roughly equivalent volume to size of cell pellet) for 8 min at 4°C. Supernatant was collected by centrifugation at max speed (13.2k rpm) for 5 min at 4°C. Total protein concentration was measured using the Bradford method with the Bio-Rad Protein Assay (#500-0006) according to the manufacturer's protocol. A total of 10 µg total protein from crude extracts was analyzed by 8% SDS-PAGE and transferred to PVDF membranes for western blot analysis. Rnt1p-HTP was detected with anti-protein A antibody conjugated to horseradish peroxidase (HRP) (Invitrogen PA1-26853) diluted 1:8000 in blocking solution (1× PBS-T, 5% milk) with the Pierce ECL Western Blotting Substrate (Thermo Fisher #32209) according to the manufacturer's protocol.

### Microscopy

Strains expressing GFP-tagged Rnt1p (Henras et al. 2004) were grown to mid-log phase in standard growth conditions and treated with 0.6 M or 0.1 M NaCl for 20 min before being prepared for imaging via the Stellaris RNA FISH protocol for *S. cerevisiae* with the following modifications: All centrifugation was performed at 850g for 6 min instead of 400g for 5 min, and Hoescht stain was used instead of DAPI for nuclear staining. The FISH probe used for visualization of mRNA localization was an oligo-d(T)25 labeled with Cy3 (MilliporeSigma). Microscopy was performed on a Leica DMI4000B Confocal microscope. All images were taken with identical settings and with a z-stack of six images with 1 µm steps, from which one z-slice was chosen from each stack for further processing. Image processing was then performed and equally applied to all images with the ImageJ software.

### Reproducibility and statistics

In general, most experiments were performed in triplicate with the exception of the experiments presented in Figures 2B, 4B–D,G which included only one replicate, and some specific lanes of some figures (see figure legends). The quantification of the western blot and northern blot signals for experiments shown in all figures and for all replicates is included in Supplemental Table S1, with the calculation of the average and the standard deviations. Paired t-test P-values are included in Supplemental Table S1 for the comparisons of averages corresponding to specific lanes for each experiment.

### SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

### ACKNOWLEDGMENTS

We thank Joseph Ong for help with the confocal microscope and Dr. Anne Hong-Hermesdorf for providing us access to the microscope. This work was supported by the National Institute of General Medical Sciences (R35 GM130370 to G.F.C.) and a National Research Service Award Training Grant (GM007185 to C.W. and K.B.).

Received July 2, 2021; accepted September 1, 2021.

### REFERENCES

- Carmody SR, Tran EJ, Apponi LH, Corbett AH, Wente SR. 2010. The mitogen-activated protein kinase Sit2 regulates nuclear retention of non-heat shock mRNAs during heat shock-induced stress. *Mol Cell Biol* **30**: 5168–5179. doi:10.1128/MCB.00735-10
- Catala M, Abou Elela S. 2019. Promoter-dependent nuclear RNA degradation ensures cell cycle-specific gene expression. *Commun Biol* **2**: 211. doi:10.1038/s42003-019-0441-3
- Catala M, Lamontagne B, Larose S, Ghazal G, Elela SA. 2004. Cell cycle-dependent nuclear localization of yeast RNase III is required for efficient cell division. *Mol Biol Cell* **15**: 3015–3030. doi:10.1091/mbc.e04-03-0183
- Chanfreau G. 2003. Conservation of RNase III processing pathways and specificity in hemiascomycetes. *Eukaryot Cell* **2**: 901–909. doi:10.1128/EC.2.5.901-909.2003
- Chanfreau G, Rotondo G, Legrain P, Jacquier A. 1998. Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1. *EMBO J* **17**: 3726–3737. doi:10.1093/emboj/17.13.3726
- Comeau M-A, Lafontaine DA, Abou Elela S. 2016. The catalytic efficiency of yeast ribonuclease III depends on substrate specific product release rate. *Nucleic Acids Res* **44**: 7911–7921. doi:10.1093/nar/gkw507
- Durant M, Pugh BF. 2007. NuA4-directed chromatin transactions throughout the *Saccharomyces cerevisiae* genome. *Mol Cell Biol* **27**: 5327–5335. doi:10.1128/MCB.00468-07
- Elela SA, Igel H, Ares M Jr. 1996. RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. *Cell* **85**: 115–124. doi:10.1016/S0092-8674(00)81087-9
- Fu J, Hou J, Liu L, Chen L, Wang M, Shen Y, Zhang Z, Bao X. 2013. Interplay between BDF1 and BDF2 and their roles in regulating the yeast salt stress response. *FEBS J* **280**: 1991–2001. doi:10.1111/febs.12219
- Gietz RD, Schiestl RH. 2007. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**: 31–34. doi:10.1038/nprot.2007.13
- Gilan O, Rioja I, Knezevic K, Bell MJ, Yeung MM, Harker NR, Lam EYN, Chung C, Bamborough P, Petretich M, et al. 2020. Selective targeting of BD1 and BD2 of the BET proteins in cancer and immuno-inflammation. *Science* **368**: eaaz8455. doi:10.1126/science.aaz8455
- Goldstein AL, McCusker JH. 1999. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. *Yeast* **15**: 1541–1553. doi:10.1002/(SICI)1097-0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K
- Granneman S, Kudla G, Petfalski E, Tollervy D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-

- linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci* **106**: 9613–9618. doi:10.1073/pnas.0901997106
- Hammell CM, Gross S, Zenklusen D, Heath C V, Stutz F, Moore C, Cole CN. 2002. Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol* **22**: 6441–6457. doi:10.1128/MCB.22.18.6441-6457.2002
- Haruki H, Nishikawa J, Laemmli UK. 2008. The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol Cell* **31**: 925–932. doi:10.1016/j.molcel.2008.07.020
- Henras AK, Bertrand E, Chanfreau G. 2004. A cotranscriptional model for 3'-end processing of the *Saccharomyces cerevisiae* pre-ribosomal RNA precursor. *RNA* **10**: 1572–1585. doi:10.1261/ma.7750804
- Henras AK, Sam M, Hiley SL, Wu H, Hughes TR, Feigon J, Chanfreau GF. 2005. Biochemical and genomic analysis of substrate recognition by the double-stranded RNA binding domain of yeast RNase III. *RNA (New York, NY)* **11**: 1225–1237. doi:10.1261/ma.2760705
- Hurt E, Strasser K, Segref A, Bailer S, Schlaich N, Presutti C, Tollervey D, Jansen R. 2000. Mex67p mediates nuclear export of a variety of RNA polymerase II transcripts. *J Biol Chem* **275**: 8361–8368. doi:10.1074/jbc.275.12.8361
- Izawa S, Kita T, Ikeda K, Inoue Y, Piper P, Tani T, Derby RJ, Hiraoka Y, Spector DL, Saavedra C, et al. 2008. Heat shock and ethanol stress provoke distinctly different responses in 3'-processing and nuclear export of HSP mRNA in *Saccharomyces cerevisiae*. *Biochem J* **414**: 111–119. doi:10.1042/BJ20071567
- Josling GA, Selvarajah SA, Petter M, Duffy MF. 2012. The role of bromodomain proteins in regulating gene expression. *Genes (Basel)* **3**: 320–343. doi:10.3390/genes3020320
- Kim KY, Levin DE. 2011. Mpk1 MAPK association with the paf1 complex blocks sen1-mediated premature transcription termination. *Cell* **144**: 745–756. doi:10.1016/j.cell.2011.01.034
- Kurdistani SK, Grunstein M. 2003. Histone acetylation and deacetylation in yeast. *Nat Rev Mol Cell Biol* **4**: 276–284. doi:10.1038/nrm1075
- Lawrence M, Dautjat S, Schneider R. 2016. Lateral thinking: how histone modifications regulate gene expression. *Trends Genet* **32**: 42–56. doi:10.1016/j.tig.2015.10.007
- Matangkasombut O, Buratowski RM, Swilling NW, Buratowski S. 2000. Bromodomain factor 1 corresponds to a missing piece of yeast TFIID. *Genes Dev* **14**: 951–962.
- Minvielle-Sebastia L, Beyer K, Krecic AM, Hector RE, Swanson MS, Keller W. 1998. Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. *EMBO J* **17**: 7454–7468. doi:10.1093/emboj/17.24.7454
- Morgado-Pascual JL, Rayego-Mateos S, Tejedor L, Suarez-Alvarez B, Ruiz-Ortega M. 2019. Bromodomain and extraterminal proteins as novel epigenetic targets for renal diseases. *Front Pharmacol* **10**: 1315. doi:10.3389/fphar.2019.01315
- Piper PW. 1995. The heat shock and ethanol stress responses of yeast exhibit extensive similarity and functional overlap. *FEMS Microbiol Lett* **134**: 121–127. doi:10.1111/j.1574-6968.1995.tb07925.x
- Piper PW, Talreja K, Panaretou B, Moradas-Ferreira P, Byrne K, Prækelt UM, Meacock P, Recnacq M, Boucherie H. 1994. Induction of major heat-shock proteins of *Saccharomyces cerevisiae*, including plasma membrane Hsp30, by ethanol levels above a critical threshold. *Microbiology* **140**: 3031–3038. doi:10.1099/13500872-140-11-3031
- Roy K, Chanfreau GF. 2012. The diverse functions of fungal RNase III enzymes in RNA metabolism. *Enzymes* **31**: 213–235. doi:10.1016/B978-0-12-404740-2.00010-0
- Roy K, Chanfreau G. 2014. Stress-induced nuclear RNA degradation pathways regulate yeast bromodomain factor 2 to promote cell survival. *PLoS Genet* **10**: e1004661. doi:10.1371/journal.pgen.1004661
- Roy K, Gabunilas J, Gillespie A, Ngo D, Chanfreau GF. 2016. Common genomic elements promote transcriptional and DNA replication roadblocks. *Genome Res* **26**: 1363–1375. doi:10.1101/gr.204776.116
- Saavedra C, Tuug KS, Amberg DC, Hopper AK, Cole CN. 1996. Regulation of mRNA export in response to stress in *Saccharomyces cerevisiae*. *Genes Dev* **10**: 1608–1620. doi:10.1101/gad.10.13.1608
- Saavedra CA, Hammell CM, Heath C V, Cole CN. 1997. Yeast heat shock mRNAs are exported through a distinct pathway defined by Rip1p. *Genes Dev* **11**: 2845–2856. doi:10.1101/gad.11.21.2845
- Schmid M, Olszewski P, Pelechano V, Gupta I, Steinmetz LM, Jensen TH. 2015. The nuclear polyA-binding protein Nab2p is essential for mRNA production. *Cell Rep* **12**: 128–139. doi:10.1016/j.celrep.2015.06.008
- Storici F, Resnick MA. 2006. The *delitto perfetto* approach to *in vivo* site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol* **409**: 329–345. doi:10.1016/S0076-6879(05)09019-1
- Treck T, Larson DR, Moldón A, Query CC, Singer RH. 2011. Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell* **147**: 1484–1497. doi:10.1016/j.cell.2011.11.051
- Tudek A, Schmid M, Makaras M, Barras JD, Beggs JD, Jensen TH. 2018. A nuclear export block triggers the decay of newly synthesized polyadenylated RNA. *Cell Rep* **24**: 2457–2467. doi:10.1016/j.celrep.2018.07.103
- Varela JOCS, Prækelt UM, Meacock PA, Planta RJ, Mager WH. 1995. The *Saccharomyces cerevisiae* HSP12 gene is activated by the high-osmolarity glycerol pathway and negatively regulated by protein kinase A. *Mol Cell Biol* **15**: 6232–6245. doi:10.1128/MCB.15.11.6232
- Volanakis A, Passoni M, Hector RD, Shah S, Kilchert C, Granneman S, Vasiljeva L. 2013. Spliceosome-mediated decay (SMD) regulates expression of nonintron genes in budding yeast. *Genes Dev* **27**: 2025–2038. doi:10.1101/gad.221960.113
- Wang C, Liu Y, DeMario SM, Mandric I, Gonzalez-Figueroa C, Chanfreau GF. 2020. Rrp6 moonlights in an RNA exosome-independent manner to promote cell survival and gene expression during stress. *Cell Rep* **31**: 107754. doi:10.1016/j.celrep.2020.107754
- Zander G, Hackmann A, Bender L, Becker D, Lingner T, Salinas G, Krebber H. 2016. mRNA quality control is bypassed for immediate export of stress-responsive transcripts. *Nature* **540**: 593–596. doi:10.1038/nature20572
- Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

**Chapter 5: Splicing inactivation generates hybrid mRNA-snoRNA transcripts targeted by cytoplasmic RNA decay**



# Splicing inactivation generates hybrid mRNA-snoRNA transcripts targeted by cytoplasmic RNA decay

Yanru Liu<sup>a,1,2</sup>, Samuel DeMario<sup>a,1</sup>, Kevin He<sup>a,1</sup>, Michelle R. Gibbs<sup>a</sup>, Keaton W. Barr<sup>a</sup>, and Guillaume F. Chanfreau<sup>a,b,3</sup>

Edited by Joan Steitz, Yale University, New Haven, CT; received February 10, 2022; accepted June 14, 2022

Many small nucleolar RNAs (snoRNA)s are processed from introns of host genes, but the importance of splicing for proper biogenesis and the fate of the snoRNAs is not well understood. Here, we show that inactivation of splicing factors or mutation of splicing signals leads to the accumulation of partially processed hybrid messenger RNA-snoRNA (hmsnoRNA) transcripts. hmsnoRNAs are processed to the mature 3' ends of the snoRNAs by the nuclear exosome and bound by small nucleolar ribonucleoproteins. hmsnoRNAs are unaffected by translation-coupled RNA quality-control pathways, but they are degraded by the major cytoplasmic exonuclease Xrn1p, due to their messenger RNA (mRNA)-like 5' extensions. These results show that completion of splicing is required to promote complete and accurate processing of intron-encoded snoRNAs and that splicing defects lead to degradation of hybrid mRNA-snoRNA species by cytoplasmic decay, underscoring the importance of splicing for the biogenesis of intron-encoded snoRNAs.

snoRNA | splicing | introns | exosome | exonuclease

Small nucleolar RNAs (snoRNAs) are noncoding RNAs that guide small nucleolar ribonucleoprotein (snoRNP)-mediated 2'-O-methylation or pseudouridylation of pre-ribosomal RNA precursors and other stable RNAs (1–3). These snoRNAs are classified in two major families: the C/D and H/ACA classes, which guide 2'-O-methylation and pseudouridylation of their substrates, respectively. The importance of correct snoRNA expression is underscored by the fact that defects in snoRNA metabolism are linked to multiple pathologic processes, including cancer (4, 5), Prader-Willi syndrome (6), and metabolic stress (7). snoRNAs are found in many different genomic contexts (2, 8). In mammalian genomes, many snoRNAs are present in the introns of host genes, but they can also be generated from long noncoding RNAs (lncRNAs). In plants, snoRNAs are often generated from polycistronic transcription units (8). In the budding yeast *Saccharomyces cerevisiae*, which has been used extensively to study the mechanisms of snoRNA biogenesis and processing, snoRNAs are expressed either from independently transcribed genes, polycistronic snoRNA precursors, or from introns (2, 8). The synthesis of mature snoRNAs from intronic sequences is thought to occur primarily through splicing of the host gene precursor messenger RNA (pre-mRNA). In vitro and in vivo studies have shown that splicing results in accurate processing of intron-encoded snoRNA in mammalian cells (9). The current model is that intron-encoded snoRNAs are generated by exonucleolytic trimming of the excised linear introns after completion of the splicing reaction and debranching of the lariat introns. In support of this model, inactivation of the *S. cerevisiae* debranching enzyme Dbr1p results in the accumulation of lariat intron species containing the snoRNAs (10), showing that debranching of the excised intron is critical for processing. However, the production of mature snoRNA can still occur inefficiently in the absence of the debranching enzyme (10), because random hydrolytic cleavage of the lariat intron exposes the cleaved intron to processing by exonucleases, or through cleavage by the RNase III enzyme Rnt1p (11). In addition, some studies have reported splicing-independent processing of intron-encoded snoRNAs (12), either by endonucleolytic cleavage of the precursors (11, 13) or by exonucleolytic processing (14). In *S. cerevisiae*, the RNase P endonuclease has also been proposed to initiate a processing pathway independent from splicing, by cleaving unspliced pre-mRNAs that host box C/D snoRNAs (15). Finally, snoRNAs can be found associated with stable lariats (16), and processing of introns containing multiple snoRNAs can generate sno-lncRNAs, which correspond to partially processed introns that contain two snoRNAs linked by an intronic segment (17).

Despite the known importance of splicing for the processing of intron-encoded snoRNAs, it is unclear how defects in the spliceosome machinery may impact the fate of intron-encoded snoRNAs. This is an important question, as recent work has shown that defects in 5'-end processing of independently transcribed snoRNAs can result in

## Significance

Small nucleolar RNAs (snoRNAs) mediate modifications of nucleosides within ribosomal RNAs, which are necessary for proper ribosomal function and translation. Many snoRNAs are encoded within introns of host genes, and accurate biogenesis of these small RNAs is required to produce functional snoRNAs. The work presented here shows that when the splicing reactions are inactivated, snoRNAs undergo a distinct biogenesis pathway, which leads to the production of aberrant hybrid RNAs that contain both messenger RNA (mRNA) and small RNA components of the host genes. While snoRNAs are primarily found in the nucleolus, these hybrid RNAs are degraded by the cytoplasmic mRNA degradation pathway. These results demonstrate the importance of splicing to promote accurate snoRNA processing and prevent the production of aberrant mRNA-snoRNA hybrids.

Author contributions: S.D. and G.F.C. designed research; Y.L., S.D., K.H., M.R.G., and K.W.B. performed research; S.D., K.H., and G.F.C. analyzed data; and G.F.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

<sup>1</sup>Y.L. and S.D. contributed equally to this work.

<sup>2</sup>Present address: Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853-2703.

<sup>3</sup>To whom correspondence may be addressed. Email: guillom@chem.ucla.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2202473119/-/DCSupplemental>.

Published July 25, 2022.

mis-localization of the unprocessed snoRNAs in budding yeast (18). Strikingly, very little is known about the impact of spliceosome defects on snoRNA expression. Haploinsufficiency of the core small nuclear ribonucleoprotein (snRNP) SmD3 results in a reduction in the levels of intron-encoded snoRNAs (19), but the specific molecular effects of this mutation on the biogenesis pathway of intron-encoded snoRNAs have not been investigated. In this study, we have analyzed the impact of inactivating splicing using trans- and cis-acting splicing mutants on the production of intron-encoded snoRNAs in the yeast *S. cerevisiae*. We show that inactivating splicing factors involved in different steps of the spliceosome cycle or mutating the splicing signal of a host gene result in the accumulation of aberrant, hybrid mRNA-snoRNAs species, which share some of the hallmarks of mature snoRNAs but are degraded by the cytoplasmic decay. These results highlight the importance of splicing for determining the fate of intron-encoded snoRNAs and show that incorrectly processed intron-encoded snoRNAs are degraded by the general messenger RNA (mRNA) decay pathway. Our results may also provide some insights into possible additional effects of genetic diseases that affect splicing factors.

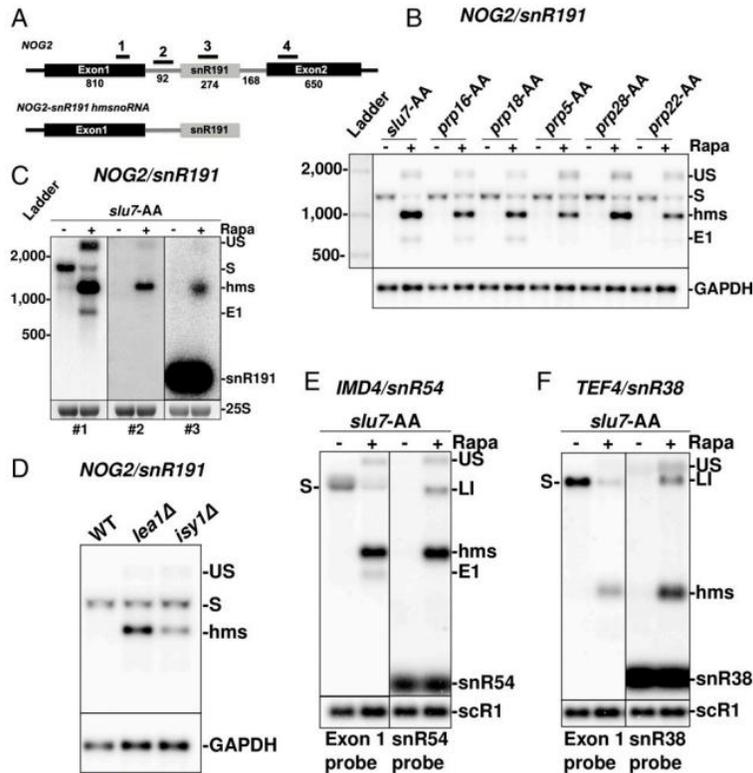
## Results

**Inactivation of Splicing Factors at Different Stages of the Cycle Results in the Accumulation of Hybrid mRNA-snoRNA Forms of *NOG2/snR191*.** To investigate the importance of the splicing reaction for the processing of intron-encoded snoRNAs, we used the anchor away (AA) technique (20), which promotes the rapid export of endogenously FRB (*FKBP12*-rapamycin binding domain)-tagged splicing factors to the cytoplasm after addition of rapamycin (20). This results in the nuclear depletion of these factors and in the inactivation of their nuclear functions in vivo. The genetic background used for these experiments alleviates the toxic effects of rapamycin and its downstream effects on gene regulation (20). To analyze the contribution of splicing factors involved at different steps of the spliceosome cycle, we used strains expressing FRB-tagged versions of Prp5p, Prp28p, Prp16p, Prp18p, Slu7p, and Prp22p. These proteins are involved in various steps of splicing (21), from spliceosome assembly (Prp5p, Prp28p) to the second catalytic step (Prp16p, Prp18p, Slu7p, Prp22p) and spliceosome recycling (Prp22p). We first analyzed expression of the *NOG2* gene, which contains the H/ACA snoRNA snR191 in its intron (22), by Northern blot (Fig. 1A). After nuclear depletion of each of these splicing factors, a probe hybridizing to the first exon of *NOG2* detected the unspliced precursors but also RNAs migrating faster than the spliced mRNAs (labeled “hms” in Fig. 1B). hmsRNAs were detected in all the AA strains treated with rapamycin but not prior to rapamycin treatment. We first hypothesized that these species might correspond to cleaved 5' exons; however, based on their estimated size (~1100 nucleotides [nt]), these species are too large to correspond to free 5' exons. In addition, these RNAs were also detected after anchoring away proteins involved prior to the first catalytic step (e.g., Prp5p, Prp28p). Furthermore, species migrating faster than the hms RNAs and whose size matched those of cleaved 5' exons (~850 nt) were detected only in strains in which second-step splicing factors are anchored away (i.e., Slu7p, Prp16p, Prp18p, Prp22p), but not in strains in which splicing factors involved prior to the first step are depleted from the nucleus (i.e., Prp5p, Prp28p). These fastest migrating species correspond to cleaved 5' exons that fail to undergo splicing because of second-step defects and are labeled “E1” in all figures.

Further mapping of the hms RNAs, using probes hybridizing to the different regions of the *NOG2-snR191* gene (Fig. 1A), showed that these species also hybridize to probes complementary to the mature snR191 snoRNA and to the intronic region preceding the snoRNA (Fig. 1C). Based on their approximate size (~1100 nt) and their hybridization patterns, we hypothesized that these RNAs correspond to hybrid mRNA-snoRNA (hmsnoRNA) containing the mRNA 5' exon, the intronic segment upstream of the snoRNA, and the entire snoRNA sequence (schematic representation is shown in Fig. 1A). The combination of estimated sizes and probe hybridization patterns was consistent with a 3' end close to that of the mature snoRNA 3' end. To test this idea, we mapped the 3' end of the *NOG2-snR191* hmsnoRNA using a modified 3'-Rapid Amplification of cDNA Ends (RACE) protocol after in vitro polyadenylation of total RNAs. Sequencing of the 3'-RACE products showed that the *snR191* hmsnoRNA 3' ends match precisely those of mature snoRNAs (SI Appendix, Fig. S1). RNAs similar in size to the hms species also accumulated in mutants carrying deletions of the nonessential genes encoding the U2 snRNP component Lea1p (23) or the splicing fidelity factor Isy1p (24) (Fig. 1D). Therefore, the production of hmsnoRNAs is caused by general splicing defects rather than by indirect effects of the AA process.

**HmsnoRNAs Can Be Detected for Most Intron-Encoded *S. cerevisiae* snoRNAs by Northern Blot and Nanopore Long-Read Sequencing.** To extend the results described above for other snoRNAs, we analyzed RNAs accumulating in the Slu7p-AA strain prior to or after rapamycin treatment for several genes expressing box C/D snoRNAs from their introns: *IMD4/snR54* (Fig. 1E), *TEF4/snR38* (Fig. 1F), and *ASC1/snR24* (SI Appendix, Fig. S2). For *IMD4/snR54* and *TEF4/snR38*, we detected the accumulation of hmsnoRNAs hybridizing to both exon 1 and snoRNA probes (Fig. 1E and F). Mapping of the *ASC1/snR24* hmsnoRNA using four probes hybridizing to different regions of *ASC1* (SI Appendix, Fig. S2 A and B) showed that these species contain the first exon of *ASC1*, the first part of the intron, and the entire snoRNA sequence that ends close to the mature 3' end (SI Appendix, Fig. S2 A and B). The estimated size and hybridization patterns of the *IMD4/snR54* and *TEF4/snR38* hmsnoRNAs (Fig. 1E and F) suggest a similar architecture. Mapping of the 3' end of the *IMD4/snR54* hmsnoRNA using the modified 3'-RACE protocol showed that the *IMD4/snR54* hmsnoRNA 3' end is identical to that of the mature *snR54* (SI Appendix, Fig. S1). Thus, splicing inactivation results in the general accumulation of hybrid mRNA-snoRNA species for all intron-encoded snoRNAs tested, which belong to both H/ACA and C/D families.

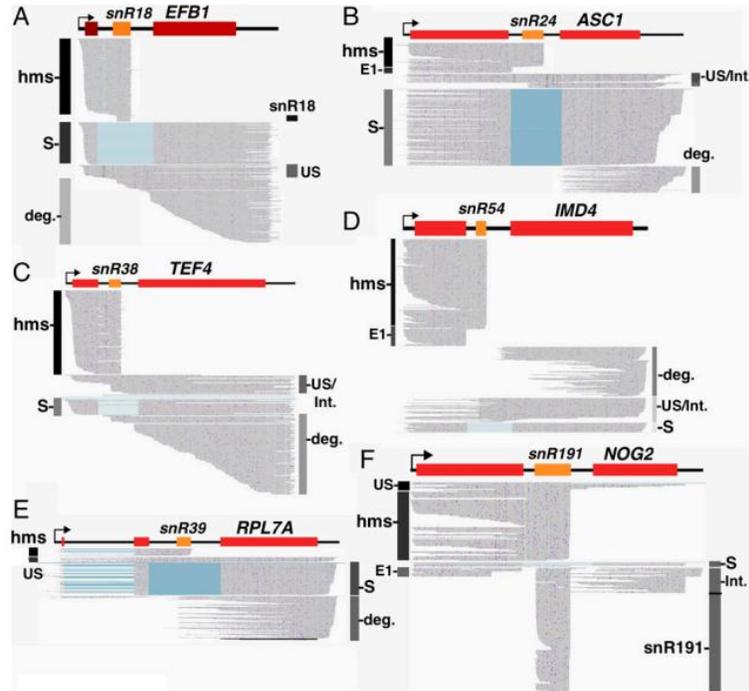
To perform a more comprehensive mapping of hmsnoRNA genome-wide, we performed a direct RNA long-read sequencing experiment on RNAs extracted from an Slu7-AA strain after addition of rapamycin, using the Oxford Nanopore platform. Prior analysis revealed that the majority of the *ASC1/snR24* hmsnoRNAs are not polyadenylated (SI Appendix, Fig. S2C), so the standard protocol to perform direct poly(A)+ RNA sequencing could not be used. We developed a protocol that used in vitro polyadenylation after ribosomal RNA (rRNA) depletion, which allowed detection of the hmsnoRNA by Nanopore sequencing. In the strain in which Slu7p was anchored away, hmsnoRNAs could be detected for *EFB1-snR18* (Fig. 2A), *ASC1-snR24* (Fig. 2B), *TEF4-snR38* (Fig. 2C), *IMD4-snR54* (Fig. 2D), *RPL7A-snR39* (Fig. 2E), *NOG2-snR191* (Fig. 2F), and *RPL7B-snR59* (SI Appendix, Fig. S3A). The architecture of hmsnoRNAs revealed by long-read sequencing matched perfectly the predictions made above



**Fig. 1.** Hybrid forms of intron-encoded snoRNAs are produced upon splicing inactivation. (A) Schematic structure of the *NOG2* gene encoding the *snR191* H/ACA snoRNA in its intron, and proposed structure of the *NOG2-snR191* hmsnoRNA. Exons are represented by black boxes and the mature snoRNA by a gray box. Intronic sequences are shown as gray lines. Numbers indicate the length, in nts, of the different exonic and intronic regions and of the snoRNA. UTR lengths are not included. Black lines with numbers 1 to 3 indicate the approximate locations of the different riboprobes used for this figure. Boxes and line lengths are not to scale. (B) Northern blot analysis of *NOG2/snR191* in strains expressing AA FRB-tagged versions of the Slu7p, Prp16p, Prp18p, Prp5p, Prp28p, and Prp22p splicing factors. Each FRB-tagged strain was grown in normal medium and then spun down and resuspended in either fresh normal medium or shifted to medium containing rapamycin (Rapa) for 1 h to promote export of the tagged splicing factor out of the nucleus. (Top) Probe 1 was used. GAPDH was used as a loading control. ; (C) Mapping of the hmsnoRNA forms of *snR191* using different riboprobes. Shown are Northern blots of *NOG2/snR191* using riboprobes 1, 2, and 3 shown in panel A of RNAs extracted from the Slu7-AA strain before or after 1 h of treatment with rapamycin. An ethidium bromide staining of the 25S rRNA is shown as a loading control. (D) Wild-type (WT) Northern blot analysis of *NOG2* in strains in WT and *lea1Δ* or *isy1Δ* deletion strains. (E) Northern blot analysis of *IMD4/snR54* in RNAs extracted from the Slu7p-AA strain before or after treatment with rapamycin for 1 h. Shown are Northern blots using probes hybridizing to the 5' exon or to the mature snoRNA. Labeling of the different species is as in B. *scr1* was used as a loading control. (F) Northern blot analysis of *TEF4/snR38* in the Slu7p-AA strain. E1, cleaved 5' exon; LI, lariar intron-exon 2 intermediate; S, spliced mRNA; US, unspliced pre-mRNA.

using Northern blot and confirmed that these species contain the 5' exon, the beginning of the intron, and the snoRNA up the mature 3' ends. We also detected unspliced precursors and cleaved 5' exons that accumulate due to the second-step splicing defect (Fig. 2). We did not detect any hmsnoRNA for the *RPS22B-snR44* gene (*SI Appendix, Fig. S3B*), likely because of the presence of an RNase III cleavage site in the first intron of the unspliced precursors, which precludes the accumulation of exon1-intron1-exon2-snR44 species when unspliced species are generated (25). Our sequencing approach produced many reads for the box H/ACA *snR191* (Fig. 2F) but failed to detect most mature box C/D snoRNA, with the exception of a few reads for *snR18* (Fig. 2A). This lack of detection may be linked to the small size of box C/D snoRNAs, which makes their mapping challenging by Oxford Nanopore sequencing. Overall, these results show that after splicing inactivation, hmsnoRNAs are generated from most intron-encoded snoRNAs in *S. cerevisiae*, and the reads obtained confirmed the general architecture of these RNAs.

**HmsnoRNAs Are Processed at Their 3' Ends by Exonucleolytic Trimming by the Exosome.** The previous results using 3'-RACE and Nanopore sequencing showed that the 3' ends of hmsnoRNAs are identical to those of mature snoRNAs, which are generated by exonucleolytic trimming by the nuclear exosome (26). To test the hypothesis that hmsnoRNAs acquire their 3' end through exosome-mediated processing, we used an AA strain in which the nuclear exosome component Rrp6p is FRB tagged. Anchoring away Rrp6p provides a more effective method of inactivating the nuclear exosome than deleting the gene encoding Rrp6p (27), possibly because exporting Rrp6p to the cytoplasm may also result in the nuclear depletion of other nuclear-exosome components. We then generated a strain in which Rrp6p and Prp18p or Slu7p could be coanchored away simultaneously. Nuclear depletion of Rrp6p resulted in an increased accumulation of both the unspliced and spliced forms of *NOG2* (Fig. 3A). This observation is consistent with previous studies that showed that the nuclear exosome actively degrades unspliced precursors and may limit the



**Fig. 2.** Detection of hmsnoRNAs by direct single-molecule Oxford Nanopore RNA sequencing. For all the panels, each sequencing read is represented by a horizontal gray line. RNA species are labeled as follows: US, unspliced pre-mRNA; S, spliced mRNA; E1, cleaved 5' exon; deg., degradation intermediates; Int., snoRNA processing intermediates. (A–F) Oxford Nanopore sequencing reads obtained from a *slu7*-FRB-tagged strain treated for 1 h with rapamycin for the following loci: (A) *EFB1*-*snR18*; (B) *ASC1*-*snR24*; (C) *TEF4*-*snR38*; (D) *IMD4*-*snR54*; *RPL7A*-*snR39*; and (F) *NOG2*-*snR191*.

production of mature species (28–30). Hybridization with a probe complementary to snR191 detected the accumulation of shortly extended forms of snR191 in the Rrp6p-AA strain (*SI Appendix, Fig. S4*), consistent with the known role of the exosome in trimming intronic snoRNA mature 3' ends (26).

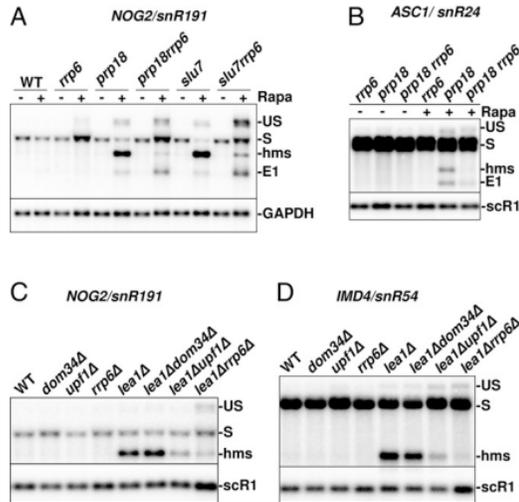
As shown previously, anchoring away Slu7p (Fig. 3*A* and *SI Appendix, Fig. S4*) or Prp18p (Fig. 3*A*), resulted in the accumulation of *NOG2*-*snR191* hmsnoRNAs. Strikingly, when Rrp6p was anchored away simultaneously with Prp18p or Slu7p, the accumulation of the hmsnoRNA decreased compared with what was observed when anchoring away these splicing factors alone (Fig. 3*A*; see also *SI Appendix, Fig. S4* for Slu7p). Similar results were obtained for *ASC1*-*snR24* (Fig. 3*B*). We observed a concomitant increase in the level of unspliced precursors (Fig. 3*A* and *SI Appendix, Fig. S4*). These results suggested a precursor-product relationship between unspliced pre-mRNAs and hmsnoRNAs, and that hmsnoRNAs are generated by exonucleolytic trimming of unspliced precursors by the nuclear exosome.

To further demonstrate that hmsnoRNAs are processed to or near the 3' ends of the mature snoRNAs by the exosome, we used strains in which the gene encoding the nonessential U2 snRNP component Lea1p is deleted, along with a deletion of the gene encoding Rrp6p. Consistent with the results obtained above, the *rrp6Δ* knockout resulted in a reduction in the accumulation of *NOG2*-*snR191* hmsnoRNA species when coupled to the *lea1Δ* mutation, accompanied by an accumulation of unspliced precursors (Fig. 3*C*). Overall, these results, combined with the analyses described above, show that hmsnoRNAs are

processed by the nuclear exosome to generate 3' ends identical to those of the corresponding mature snoRNAs.

#### HmsnoRNAs Are Not Targeted by Translation-Dependent Quality-Control Pathways.

Because hmsnoRNAs contain 5'-untranslated region (UTR) and exon 1 sequences, they exhibit mRNA-like features at their 5' ends, which suggests that they might be translated. However, they lack a poly(A) tail, which may limit translation efficiency. If hmsnoRNAs are translated, translation would stop shortly after the end of the 5'-exon sequence because of the occurrence of premature stop codons in the intronic sequences, which may trigger nonsense-mediated decay (NMD). NMD could potentially affect hmsnoRNAs, since NMD can occur in the absence of a poly(A) tail in yeast (31). Alternatively, the presence of secondary structures in the snoRNAs or the binding by snoRNPs (see below) may block ribosome progression and trigger No-Go Decay (32). To investigate a potential targeting of hmsnoRNAs by these two RNA surveillance pathways, we knocked out the genes encoding the NMD factor Upf1p or the No-Go Decay factor Dom34p in the *lea1Δ* mutant and analyzed *NOG2*-*snR191* and *IMD4*-*snR54* by Northern blot (Fig. 3*C* and *D*). Eliminating Upf1p or Dom34p did not increase the levels of hmsnoRNAs in the *lea1Δ* mutant (Fig. 3*C* and *D*). Instead, the *upf1Δ* knockout resulted in a decrease of the level of hmsnoRNAs, while eliminating Dom34p showed no effect. As opposed to what was observed in the *lea1Δrrp6Δ* mutant, the decreased accumulation of the *NOG2*-*snR191* hmsnoRNA detected in the *lea1Δupf1Δ* mutant did not



**Fig. 3.** hmsnoRNAs acquire their 3' ends through processing by the nuclear exosome but are mostly unaffected by cytoplasmic quality-control pathways. (A) Northern blot analysis of *NOG2/snR191* in strains expressing FRB-tagged, AA versions of Rrp6p, Prp18p, Slu7p, or double AA versions. Strains in the name of the gene italicized indicate that the corresponding protein was FRB tagged (e.g., *rrp6 = rrp6-FRB*). The membrane was hybridized with a probe complementary to the exon 1 of *NOG2*. GAPDH was used as a loading control. E1, cleaved exon 1; Rapa, rapamycin; S, spliced mRNA; US, unspliced pre-mRNA; WT, wild type. (B) Northern blot analysis of *ASC1/snR191* in strains expressing FRB-tagged, AA versions of Rrp6p, Prp18p, or both Prp18p and Rrp6p. The probe used was complementary to the exon 1 of *ASC1*. ScR1 was used as a loading control. Labeling of the species is as in A. (C) Northern blot analysis of *NOG2/snR191* in strains carrying viable deletions of the genes encoding the splicing factor *Lea1p* and/or the RNA degradation or processing factors *Dom34p*, *Upf1p*, or *Rrp6p*. The probe used is complementary to the exon 1 of *NOG2*. ScR1 was used as a loading control. Labeling of the species is as in A. (D) Analysis of *IMD4/snR54* as described for *NOG2/snR191* in C.

correlate with an increased accumulation of the unspliced precursors for snR191 (Fig. 3C), suggesting that Upf1p is not involved in the conversion of unspliced precursors into hmsnoRNAs. We detected some unspliced precursor accumulation for *IMD4* in the *lea1Δupf1Δ* mutant (Fig. 3D), but this was also detected in the *upf1Δ* mutant. While we do not fully understand the molecular basis for the decreased accumulation of hmsnoRNAs in the *lea1Δupf1Δ* mutant, experiments described below using splicing signals mutations suggest that this might be the result of indirect effects. Overall, we conclude from these experiments that hmsnoRNAs are not targeted by translation-mediated RNA quality-control pathways.

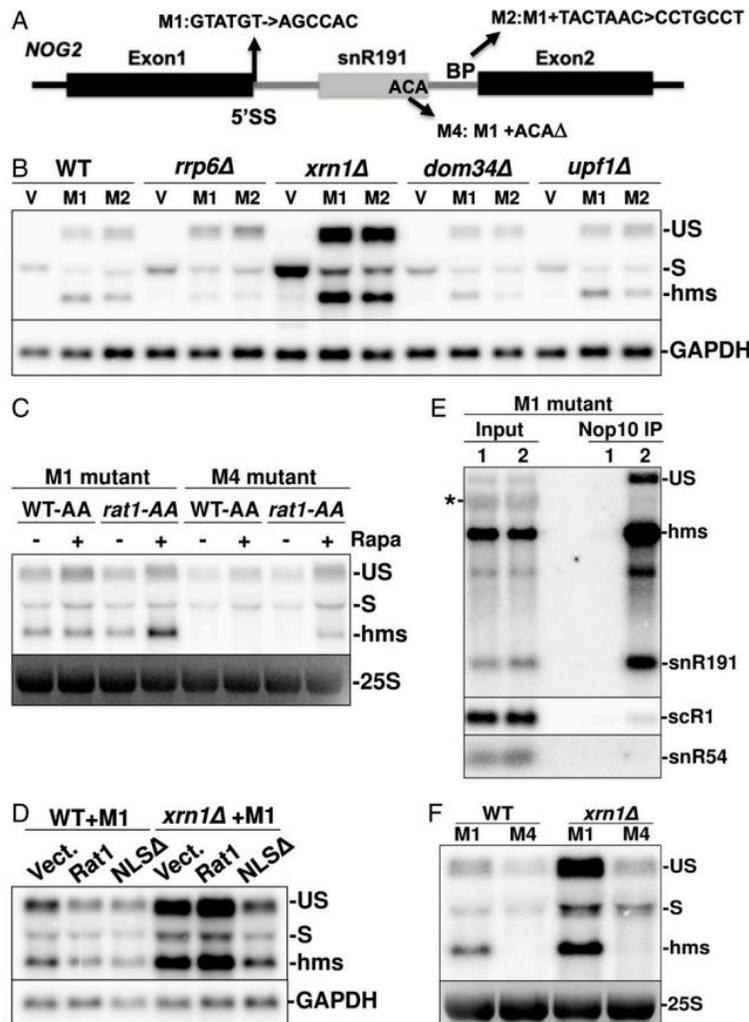
**hmsnoRNAs Can Be Generated by a 5' Splice-Site Mutation and Are Degraded by the Cytoplasmic Decay Pathway Involving Xrn1p.** The previous results showed that inactivation of splicing factors resulted in the accumulation of hybrid mRNA-snoRNA species. However, we could not rule out that anchoring away splicing factors or deleting genes encoding nonessential splicing factors may result in indirect effects that may hamper the interpretation of our results. To alleviate these concerns, we used a plasmid-based system to introduce splice-site mutations that would inhibit splicing and circumvent any general splicing defects. We generated a construct expressing the *NOG2-snR191* gene from a centromeric plasmid, which expressed *NOG2* at levels that exceeded the RNAs expressed from the endogenous *NOG2* gene

(SI Appendix, Fig. S5). We did not detect any hmsnoRNA or any other aberrant species from the plasmid-borne wild-type version of *NOG2*, showing that it is fully and accurately processed (SI Appendix, Fig. S5). We then generated two mutants designed to inactivate splicing of the plasmid-borne *NOG2* transcripts: M1 is a mutation of the 5' splice site (SS), and M2 is a double mutation that combines M1 with a branch point (BP) mutation (Fig. 4A). Since the plasmid-borne versions are expressed at higher levels than the endogenous *NOG2* transcripts, and because *NOG2* is an essential gene, we used these plasmids in strains that also expressed the endogenous *NOG2* gene, which facilitated the genetic analysis. Transcripts expressed from the M1 and M2 plasmids accumulated in wild-type strains mostly as hmsnoRNAs and, to a lesser extent, as unspliced transcripts (Fig. 4B). The temperature used to grow strains had some influence on the level of accumulation of hmsnoRNAs derived from these mutants, as hmsnoRNAs were more abundant when cells were grown at 20 °C compared with 30 °C (SI Appendix, Fig. S6).

In the *rrp6Δ* mutant, the levels of hmsnoRNAs expressed from the M1 or M2 plasmids decreased with an accumulation of unspliced precursors, consistent with our previous conclusion that the nuclear exosome processes the 3' end of hmsnoRNAs. The *dom34Δ* and *upf1Δ* mutations had no major effect on the accumulation of hmsnoRNAs expressed from these plasmids, in contrast to the result described above, which showed a decreased accumulation of hmsnoRNAs in the *lea1Δupf1Δ* mutant (Fig. 3D). Since inactivating Upf1p does not impact the level of hmsnoRNAs generated by splicing-signal mutations, we interpret the decreased accumulation of hmsnoRNA previously observed in the *lea1Δupf1Δ* mutant as the result of indirect effects. This interpretation is supported by previous work showing that mutations that reduce splicing efficiency and inactivate NMD result in synthetic growth defects (33), which may indirectly impair processes required for the accumulation of hmsnoRNAs.

In contrast to all other RNA degradation mutants analyzed previously, a large increase of all *NOG2* RNA species was detected in the *xrn1Δ* mutant deficient for the major cytoplasmic 5'-3' exonuclease (34) (Fig. 4B). While the increased accumulation of spliced and unspliced species was expected, the increase in abundance of the hmsnoRNAs strongly suggests that these species are degraded in the cytoplasm by the general mRNA decay pathway, which relies primarily on Xrn1p. As hmsnoRNAs contain the 5' ends and exon 1 sequences of the *NOG2* gene, they may contain a 7mG-cap similar to that of mRNAs, which would subject them to the general decay pathway for cytoplasmic mRNAs (35). In order to show that hmsnoRNAs are degraded by the general cytoplasmic decay pathway, which requires decapping, we analyzed hmsnoRNA accumulation resulting from the M1 mutation in a strain lacking the decapping enzyme Dcp2p. As observed previously for the *xrn1Δ* mutant, inactivation of Dcp2p resulted in an increased accumulation of the hmsnoRNAs (SI Appendix, Fig. S7). This result is consistent with the idea that hmsnoRNAs are degraded by the major cytoplasmic decay pathway, which involves Dcp1/2-mediated decapping and 5'-3' degradation by Xrn1p.

**hmsnoRNAs Are Affected by Delocalization of Rat1p in the Cytoplasm.** To test if hmsnoRNAs are affected by the nuclear 5'-3' exonuclease Rat1p (also known as Xrn2p), we inactivated Rat1p by AA and assessed the accumulation of hmsnoRNAs from the M1 construct (Fig. 4C). This analysis revealed higher hmsnoRNAs levels in conditions where Rat1p is depleted from the nucleus, suggesting that a fraction of hmsnoRNAs is nuclear. It is also possible that the inactivation of Rat1p may limit



**Fig. 4.** hmsnoRNAs generated by SS mutations are degraded by Xrn1p and Rat1p and share some of the features of mature snoRNPs. (A) Mutations introduced in the plasmid expressing *NOG2-snrR191*. (B) *NOG2-snrR191* hmsnoRNAs generated by splicing signal mutations are degraded by Xrn1p but unaffected by cytoplasmic quality-control pathways. Wild-type (WT) or the indicated deletion mutants were transformed with the pUG35 vector (V) or the pUG35 plasmids expressing mutants M1 or M2, and RNAs extracted from these strains were analyzed by Northern blot using a probe hybridizing to the exon 1 of *NOG2*. GAPDH was used as a loading control. RNA species are labeled as follows: E1, cleaved exon 1; S, spliced mRNA; US, unspliced pre-mRNA. (C) Impact of Rat1p inactivation on hmsnoRNA levels expressed from the M1 and M4 plasmids. A WT strain or a Rat1-FRB AA strain transformed with the M1 or M4 expression plasmids were grown in medium in the absence or presence of rapamycin (Rapa) for 1 h and analyzed by Northern blot as described for B. 25S rRNA was used as a loading control. (D) Impact of Rat1p delocalization on hmsnoRNA levels expressed from the M1 plasmid. A WT strain or an *xrn1Δ* mutant strain transformed with the M1 expression plasmid were transformed with an empty vector (vect.), a plasmid expressing WT Rat1p (Rat1), or a version of Rat1p with a mutated nuclear localization signal (NLSΔ). RNAs were analyzed by Northern blot as described for B. (E) *NOG2-snrR191* hmsnoRNAs generated by the M1 mutation are bound by the H/ACA snoRNP Nop10p. Shown are Northern blots of RNAs extracted from cell extracts (Input) prepared from untagged wild strain (lane 1) or ZZ-tagged Nop10 strain (lane 2) and of RNAs extracted after incubation of whole-cell extracts from the same strains with IgG beads and washing (see *Materials and Methods*). Membranes corresponding to independent purification experiments were hybridized to a probe complementary to the exon 1 of *NOG2* and to *snR191* (Top) or to probes hybridizing to *scR1* or *snR54* (Bottom). \*Cross-hybridization of the probe to a ribosomal RNA. (F) Destabilization of hmsnoRNAs caused by the deletion of the ACA box of plasmid-borne *NOG2-snrR191*s cannot be rescued by inactivation of Xrn1p. Shown is a Northern blot analysis using a probe hybridizing to the exon 1 of *NOG2* of RNAs extracted from WT or *xrn1Δ* mutants transformed with plasmids expressing the M1 or M4 constructs, as in A.

degradation of unspliced precursors as shown previously (25, 28), which could then be converted into hmsnoRNAs and result in higher hmsnoRNA levels.

To assess whether Rat1p could degrade hmsnoRNAs in the cytoplasm, we used plasmids expressing wild-type Rat1p or a

mutant lacking its nuclear localization signal (NLSΔ), which was shown previously to be able to rescue Xrn1p activity (36). The expression of these constructs had little effect in a wild-type context (Fig. 4D). By contrast, expressing the *rat1-NLSΔ* mutant in a strain lacking Xrn1p resulted in a strong decrease

of hmsnoRNA and unspliced and mature mRNAs levels, while expressing a wild-type version of Rat1p had little impact. This result shows that a version of Rat1p, which is delocalized in the cytoplasm, can destabilize hmsnoRNAs in the absence of Xrn1p, providing further evidence that hmsnoRNAs are primarily degraded through cytoplasmic 5'-3' decay.

#### **hmsnoRNAs Share Some of the Hallmarks of Mature snoRNAs.**

The plasmid expressing *NOG2* with a mutated 5'-SS provided us a suitable system to investigate if hmsnoRNAs share some of the features of mature snoRNPs. We first asked if hmsnoRNAs are bound by a snoRNP protein found in mature snoRNPs. We performed immunoprecipitation (IP) of RNAs in a strain expressing the M1 mutant and a ZZ-tagged version of Nop10p (37), a core component of H/ACA snoRNPs. A control strain was included which did not express tagged Nop10p. Northern blot analysis of RNAs purified on immunoglobulin G (IgG) beads (which bind the ZZ-tagged Nop10p) showed that hmsnoRNAs copurified with Nop10p, as did the mature forms of *snR191* (Fig. 4E, IP lane 2). By contrast, non-H/ACA RNAs such as *snR54* or *scR1* were not efficiently purified (Fig. 4E and SI Appendix, Fig. S8), and no RNAs were retained on IgG beads in the strain that did not express tagged Nop10p (Fig. 4E, IP lane 1), showing the specificity of this immunoprecipitation procedure. The ratio of signals for the RNAs found in the Nop10p immunoprecipitates compared with the input showed very similar values for the mature *snR191* and the hmsnoRNAs (SI Appendix, Fig. S8), suggesting that hmsnoRNAs are quantitatively bound by Nop10p. Interestingly, unspliced precursors were also immunoprecipitated by Nop10p (Fig. 4E). This is consistent with prior work showing cotranscriptional assembly of H/ACA snoRNPs on an intron-encoded H/ACA snoRNA in *S. cerevisiae* (38).

If hmsnoRNAs are bound by snoRNPs and stabilized by their binding, one would expect that mutation of snoRNA sequence elements that promote snoRNP assembly would result in a decrease of a steady-state levels of hmsnoRNAs, as described previously for mature H/ACA snoRNAs (1). To test this hypothesis, we generated a variant of the M1 mutant in which the ACA box of *snR191* is deleted (M4 mutant; Fig. 4A). This mutation resulted in the complete destabilization of the hmsnoRNA species, which were no longer detectable (Fig. 4 C and F). Strikingly, anchoring away Rat1p from the nucleus partially restored the accumulation of hmsnoRNAs expressed from this construct (Fig. 4C). This result shows that defective assembly of snoRNPs due to the ACA-box deletion triggers degradation of the hmsnoRNAs by nuclear 5'-3' decay. By contrast, deletion of Xrn1p did not result in any rescue of hmsnoRNA levels from the M4 mutant, as opposed to what was observed for the hmsnoRNAs generated from the M1 construct (Fig. 4F). These results show that hmsnoRNAs that are not properly assembled into snoRNPs are primarily degraded by Rat1p in the nucleus and are unaffected by cytoplasmic turnover.

#### **Discussion**

In this work, we show that inactivation of splicing by impairing splicing factors or mutating splicing signals leads to the production of hybrid mRNA-snoRNA species for both box C/D or H/ACA snoRNAs. Extensive mapping of these species by Northern blot and Oxford Nanopore long-read sequencing showed that hmsnoRNA have extended 5' ends containing mRNAs sequences that include 5'-UTR and exon 1 and that their 3' ends match those of mature snoRNAs. This chimeric

architecture is reminiscent of that of sno-lncRNAs (17), but hmsnoRNAs contain mRNA-like features instead of an snoRNA at their 5' ends. The data we have presented suggest a general pathway for the production and degradation of hmsnoRNAs (SI Appendix, Fig. S9). Splicing inactivation results in accumulation of unspliced pre-mRNAs, which are bound by at least a subpopulation of snoRNPs (e.g., Nop10p for the *NOG2-snR191* hmsnoRNA) and then trimmed to or near the mature 3' end of the snoRNA sequence by the nuclear exosome (SI Appendix, Fig. S9). Early binding of snoRNPs is consistent with findings of prior studies that have shown cotranscriptional assembly of snoRNP components (38–41). This assembly can occur in a splicing-independent manner (41), which explains why Nop10p can bind to hmsnoRNAs produced when splicing is inactivated (Fig. 4E) and why deletion of the ACA box of an *NOG2-snR191* gene containing a 5'-SS mutant significantly destabilizes the RNAs generated from this construct and triggers nuclear decay by Rat1p (Fig. 4C). After nuclear 3'-end processing by the exosome, hmsnoRNAs are exported to the cytoplasm, where they can be degraded by the general 5'-3' decay pathway that involves Dcp2p and Xrn1p (SI Appendix, Fig. S9). Further evidence for cytoplasmic localization of hmsnoRNAs is provided by the fact that a version of Rat1p mis-localized in the cytoplasm can reduce hmsnoRNA levels in a strain lacking Xrn1p (Fig. 4D). Despite degradation by these 5'-3' decay pathways, hmsnoRNA can accumulate to relatively high levels, from 10 to 64% relative to the mature snoRNA levels, depending on the snoRNA (SI Appendix, Table S1).

We found that inactivation of splicing factors involved at different steps of the spliceosome cycle results in the accumulation of hmsnoRNAs. This observation might be unexpected for factors involved in the second catalytic step such as Slu7p or Prp18p, as anchoring away these factors is expected to produce mostly lariat intermediates and cleaved 5' exons but not unspliced precursors, which we showed are converted into hmsnoRNAs by the nuclear exosome. However, a recent study offers an explanation to this conundrum, as inactivation of late splicing factors can result in unspliced precursors accumulation in vivo (42), possibly because of recycling defects. Our work sheds light on previous results that described the accumulation of RNAs similar to hmsnoRNAs upon mutation of the BP sequence of the host intron of *snR18* (previously called U18) (43). The result reported in this previous study is reminiscent of the effect detected when mutating the 5' SS and the BP of *NOG2*. In the previous study, the 5'-extended *snR18* species were interpreted as intermediates in the processing pathway (43). However, there was no evidence provided that these species could be converted into functional mature products, and based on our observations, it is more likely that they correspond to dead-end products similar to hmsnoRNAs and which are defective byproducts of splicing inactivation.

The accumulation of RNAs similar to hmsnoRNAs has also been reported when the precursor transfer RNA processing enzyme RNase P is inactivated by a thermosensitive mutation (15). These species were considered intermediates in the pathway, and their accumulation was interpreted as evidence for a direct involvement of RNase P in the processing pathway of intron-encoded snoRNAs. However, RNase P-mediated cleavage intermediates could not be detected in vivo in this study (15); furthermore, lariat introns containing snoRNAs accumulate at very high levels in a debranching enzyme mutant (10), which argues against a major cleavage pathway of intronic sequences by RNase P. We propose, instead, that the accumulation of hmsnoRNA-like RNAs in this RNase P thermosensitive mutant is, in fact, due to an indirect inhibition of splicing, as

work published later by the same group showed that this mutation results in the accumulation of unspliced pre-mRNA precursors in vivo (44), which might be converted into hmsnoRNAs.

Splicing activity has been shown to be inhibited during stress or nonstandard growth conditions (45), which suggests that hmsnoRNAs could be naturally produced in such conditions without experimental interference with the splicing process. To assess whether hmsnoRNAs might be produced in growth conditions that reduce splicing efficiency, we analyzed *NOG2* expression in stationary phase, heat-shock conditions or treatment with rapamycin (in a wild-type strain that contains a functional TOR pathway, unlike the AA strains described in *Results*). None of these conditions resulted in the accumulation of hmsnoRNAs (*SI Appendix*, Fig. S10). However, the same conditions are also known to generally reduce ribosome biogenesis and snoRNP assembly. Therefore, the absence of detection of these species in these conditions is difficult to interpret, as it might be due to conditions that globally repress ribosome biogenesis and prevent assembly of snoRNPs on the hmsnoRNA species.

The results presented here provide a general framework that underscores the importance of the splicing process for the biogenesis of snoRNAs. In the absence of splicing, not only are intron-encoded snoRNAs not processed properly but the RNAs species that are improperly generated are subject to a degradation pathway similar to that which targets mRNAs in the cytoplasm (*SI Appendix*, Fig. S9). A parallel model was proposed to underscore the importance of 5'-end processing for independently transcribed snoRNAs in *S. cerevisiae* (18). In the absence of cotranscriptional cleavage of snoRNA precursors by the endonuclease Rnt1p, unprocessed snoRNAs accumulate in the cytoplasm and are not functional (18). Overall, the results described here combined with those reported by Grzechnik et al. (18) converge to establish a unified model for the importance of RNA processing for the fate of snoRNAs. Regardless of the precise mode of expression and of the nature of the transcription units that produce snoRNAs, RNA processing of snoRNA precursors serves two major purposes: these reactions not only remove flanking sequences, they also dictate the proper fate of snoRNP particles and their mode of degradation. In the case of intron-encoded snoRNAs, the work described here establishes an additional functional role for splicing reactions beyond simply removing intervening sequences of mRNAs. Finally, these

data may shed some light on the molecular effects of mutations of splicing factors, which can result in a variety of human diseases. While the impact of these mutations on mRNA splicing and alternative splicing patterns has been investigated, their effect on the production of intron-encoded snoRNAs has not been well characterized. It is possible that some of the deleterious effects of these mutations that are causal to disease may be linked to the defective processing of intron-encoded snoRNAs and the production of hmsnoRNA-like species, which may be detrimental to RNA metabolism.

## Materials and Methods

Yeast strain construction and manipulation, RNA extraction, and Northern blot analysis using riboprobes were performed as described by Wang et al. (46). The list of strains, plasmids, and oligonucleotides used is provided in the *SI Appendix*. Wild-type and knockout yeast strains are from the BY4742 genetic background (47), and knockout strains were obtained from the systematic knockout collection (48). Strains used for the AA experiments are from the HHY168 genetic background (20). The detailed protocol used for direct RNA sequencing is described in detail in the *SI Appendix* and sequencing data are available at the National Center for Biotechnology Information as BioProject ID: PRJNA827814. For the expression of *NOG2* mutants, the wild-type *NOG2* gene was cloned in pUG35 (49) and expressed under the control of its own promoter to create pCL1. The M1, M2, and M4 mutants were created by site-directed mutagenesis from pCL1. For IP of ZZ-tagged Nop10p, we used strains transformed with the pFH35 plasmid that expresses ZZ-tagged Nop10p (37). RNA IP was performed as described in the *SI Appendix*.

**Data Availability.** Sequencing data have been deposited with the National Center for Biotechnology Information BioProject (PRJNA827814). All study data are included in the article and/or supporting information.

**ACKNOWLEDGMENTS.** We thank Aidan Boyne for technical help, Anthony Henras for sharing the Nop10-ZZ tagged expression plasmid and for helpful discussions, and Jeff Coller and Arlen Johnson for sharing strains and plasmids. This work was supported by the National Institute of General Medical Sciences (NIGMS; Grant R35 GM130370 to G.F.C.) M.R.G. was supported by NIGMS grant IRACDA 2K12 GM106996-06.

Author affiliations: <sup>a</sup>Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095; and <sup>b</sup>Molecular Biology Institute, University of California, Los Angeles, CA 90095

1. A. G. Balakin, L. Smith, M. J. Fournier, The RNA world of the nucleolus: Two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823–834 (1996).
2. J. Kufel, P. Grzechnik, Small nucleolar RNAs tell a different tale. *Trends Genet.* **35**, 104–117 (2019).
3. T. Kiss, Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* **20**, 3617–3622 (2001).
4. J. Liang et al., Small nucleolar RNAs: Insight into their function in cancer. *Front. Oncol.* **9**, 587 (2019).
5. L. Faucher-Giguère, et al., High-grade ovarian cancer associated H/ACA snoRNAs promote cancer cell proliferation and survival. *NAR Cancer* **4**, zcab050 (2022).
6. J. Cavallé, Box C/D small nucleolar RNA genes and the Prader-Willi syndrome: A complex interplay. *Wiley Interdiscip. Rev. RNA* **8**, e1417 (2017).
7. J. E. Schaffer, Death by lipids: The role of small nucleolar RNAs in metabolic stress. *J. Biol. Chem.* **295**, 8628–8635 (2020).
8. J. W. Brown, D. F. Marshall, M. Echeverria, Intronic noncoding RNAs and splicing. *Trends Plant Sci.* **13**, 335–342 (2008).
9. T. Kiss, W. Filipowicz, Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev.* **9**, 1411–1424 (1995).
10. S. L. Ooi, D. A. Samarsky, M. J. Fournier, J. D. Boeke, Intronic snoRNA biosynthesis in *Saccharomyces cerevisiae* depends on the lariat-debranching enzyme: Intron length effects and activity of a precursor snoRNA. *RNA* **4**, 1096–1110 (1998).
11. G. Ghazal, D. Ge, J. Genais-Bird, J. Gagnon, S. Abou Elela, Genome-wide prediction and analysis of yeast RNase III-dependent snoRNA processing signals. *Mol. Cell. Biol.* **25**, 2981–2994 (2005).
12. T. Hirose, M. D. Shu, J. A. Steitz, Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol. Cell* **12**, 113–123 (2003).
13. E. Caffarelli, M. Arese, B. Santoro, P. Fragapane, I. Bozoni, In vitro study of processing of the intron-encoded U16 small nucleolar RNA in *Xenopus laevis*. *Mol. Cell. Biol.* **14**, 2966–2974 (1994).
14. F. Ceconi, P. Mariotti, F. Amaldi, The *Xenopus* intron-encoded U17 snoRNA is produced by exonucleolytic processing of its precursor in oocytes. *Nucleic Acids Res.* **23**, 4670–4676 (1995).
15. D. J. Coughlin, J. A. Pleiss, S. C. Walker, G. B. Whitworth, D. R. Engelke, Genome-wide search for yeast RNase P substrates reveals role in maturation of intron-encoded box C/D small nucleolar RNAs. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12218–12223 (2008).
16. G. J. S. Talross, S. Deryusheva, J. G. Gall, Stable lariats bearing a snoRNA (slb-snoRNA) in eukaryotic cells: A level of regulation for guide RNAs. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2114156118 (2021).
17. Q. F. Yin et al., Long noncoding RNAs with snoRNA ends. *Mol. Cell* **48**, 219–230 (2012).
18. P. Grzechnik, et al., Nuclear fate of yeast snoRNA is determined by co-transcriptional Rnt1 cleavage. *Nat. Commun.* **9**, 1–14 (2018).
19. B. S. Scruggs, C. I. Michel, D. S. Ory, J. E. Schaffer, Smd3 regulates intronic noncoding RNA biogenesis. *Mol. Cell. Biol.* **32**, 4092–4103 (2012).
20. H. Haruki, J. Nishikawa, U. K. Laemmli, The anchor-away technique: Rapid, conditional establishment of yeast mutant phenotypes. *Mol. Cell* **31**, 925–932 (2008).
21. M. C. Wahl, C. L. Will, R. Lührmann, The spliceosome: Design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
22. G. Badis, M. Fromont-Racine, A. Jacquier, A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast. *RNA* **9**, 771–779 (2003).
23. F. Casparly, B. Séraphin, The yeast U2A/U2B complex is required for pre-spliceosome formation. *EMBO J.* **17**, 6348–6358 (1998).
24. T. Villa, C. Guthrie, The Isy1p component of the NineTeen complex interacts with the ATPase Prp16p to regulate the fidelity of pre-mRNA splicing. *Genes Dev.* **19**, 1894–1904 (2005).
25. M. Danin-Kreiselman, C. Y. Lee, G. Chanfreau, RNase III-mediated degradation of unspliced pre-mRNAs and lariat introns. *Mol. Cell* **11**, 1279–1289 (2003).
26. C. Allmang et al., Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**, 5399–5410 (1999).
27. K. Roy, J. Gabunilas, A. Gillespie, D. Ngo, G. F. Chanfreau, Common genomic elements promote transcriptional and DNA replication roadblocks. *Genome Res.* **26**, 1363–1375 (2016).

28. C. Bousquet-Antonelli, C. Presutti, D. Tollervey, Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102**, 765–775 (2000).
29. R. K. Gudipati *et al.*, Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol. Cell* **48**, 409–421 (2012).
30. S. Sayani, G. F. Chanfreau, Sequential RNA degradation pathways provide a fail-safe mechanism to limit the accumulation of unspliced transcripts in *Saccharomyces cerevisiae*. *RNA* **18**, 1563–1572 (2012).
31. S. Meaux, A. van Hoof, K. E. Baker, Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail. *Mol. Cell* **29**, 134–140 (2008).
32. M. K. Doma, R. Parker, Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561–564 (2006).
33. T. Kawashima, M. Pellegrini, G. F. Chanfreau, Nonsense-mediated mRNA decay mutes the splicing defects of spliceosome component mutations. *RNA* **15**, 2236–2247 (2009).
34. S. Geisler, J. Collier, XRN1: A major 5' to 3' exonuclease in eukaryotic cells. *Enzymes* **31**, 97–114 (2012).
35. J. S. Mugridge, J. Collier, J. D. Gross, Structural and molecular mechanisms for the control of eukaryotic 5'-3' mRNA decay. *Nat. Struct. Mol. Biol.* **25**, 1077–1085 (2018).
36. A. W. Johnson, Rat1p and Xrn1p are functionally interchangeable exonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Mol. Cell. Biol.* **17**, 6122–6130 (1997).
37. C. Dez *et al.*, Stable expression in yeast of the mature form of human telomerase RNA depends on its association with the box H/ACA small nucleolar RNP proteins Cbf5p, Nhp2p and Nop10p. *Nucleic Acids Res.* **29**, 598–603 (2001).
38. P. K. Yang *et al.*, Cotranscriptional recruitment of the pseudouridyltransferase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. *Mol. Cell. Biol.* **25**, 3295–3304 (2005).
39. M. Ballarino, M. Morlando, F. Pagano, A. Fatica, I. Bozzoni, The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **25**, 5396–5403 (2005).
40. X. Darzacq *et al.*, Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J. Cell Biol.* **173**, 207–218 (2006).
41. P. Richard, A. M. Kiss, X. Darzacq, T. Kiss, Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Mol. Cell. Biol.* **26**, 2540–2549 (2006).
42. G. I. Mendoza-Ochoa, J. D. Barrass, I. E. Maudlin, J. D. Beggs, Blocking late stages of splicing quickly limits pre-spliceosome assembly in vivo. *RNA Biol.* **16**, 1775–1784 (2019).
43. T. Villa, F. Ceradini, I. Bozzoni, Identification of a novel element required for processing of intron-encoded box C/D small nucleolar RNAs in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **20**, 1311–1320 (2000).
44. M. C. Marvin *et al.*, Accumulation of noncoding RNA due to an RNase P defect in *Saccharomyces cerevisiae*. *RNA* **17**, 1441–1450 (2011).
45. J. A. Pleiss, G. B. Whitworth, M. Bergkessel, C. Guthrie, Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol. Cell* **27**, 928–937 (2007).
46. C. Wang *et al.*, Rrp6 moonlights in an RNA exosome-independent manner to promote cell survival and gene expression during stress. *Cell Rep.* **31**, 107754 (2020).
47. C. B. Brachmann *et al.*, Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**, 115–132 (1998).
48. E. A. Winzeler *et al.*, Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
49. R. K. Niedenthal, L. Riles, M. Johnston, J. H. Hegemann, Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast* **12**, 773–786 (1996).

## 5.2 Ongoing work and updates

One major point made in this work was that our multiple attempts to find a biological condition in which these hmsnoRNAs naturally accumulate independently from splicing inactivation were unsuccessful. Additionally, our results at the time suggested that the hmsnoRNA products were not polyadenylated, based on a northern blot analysis of the ASC1/snR24 in poly(A)<sup>+</sup> RNAs. However, recent analysis of subsequent Oxford Nanopore RNA sequencing data has shed more light on both of these subjects.

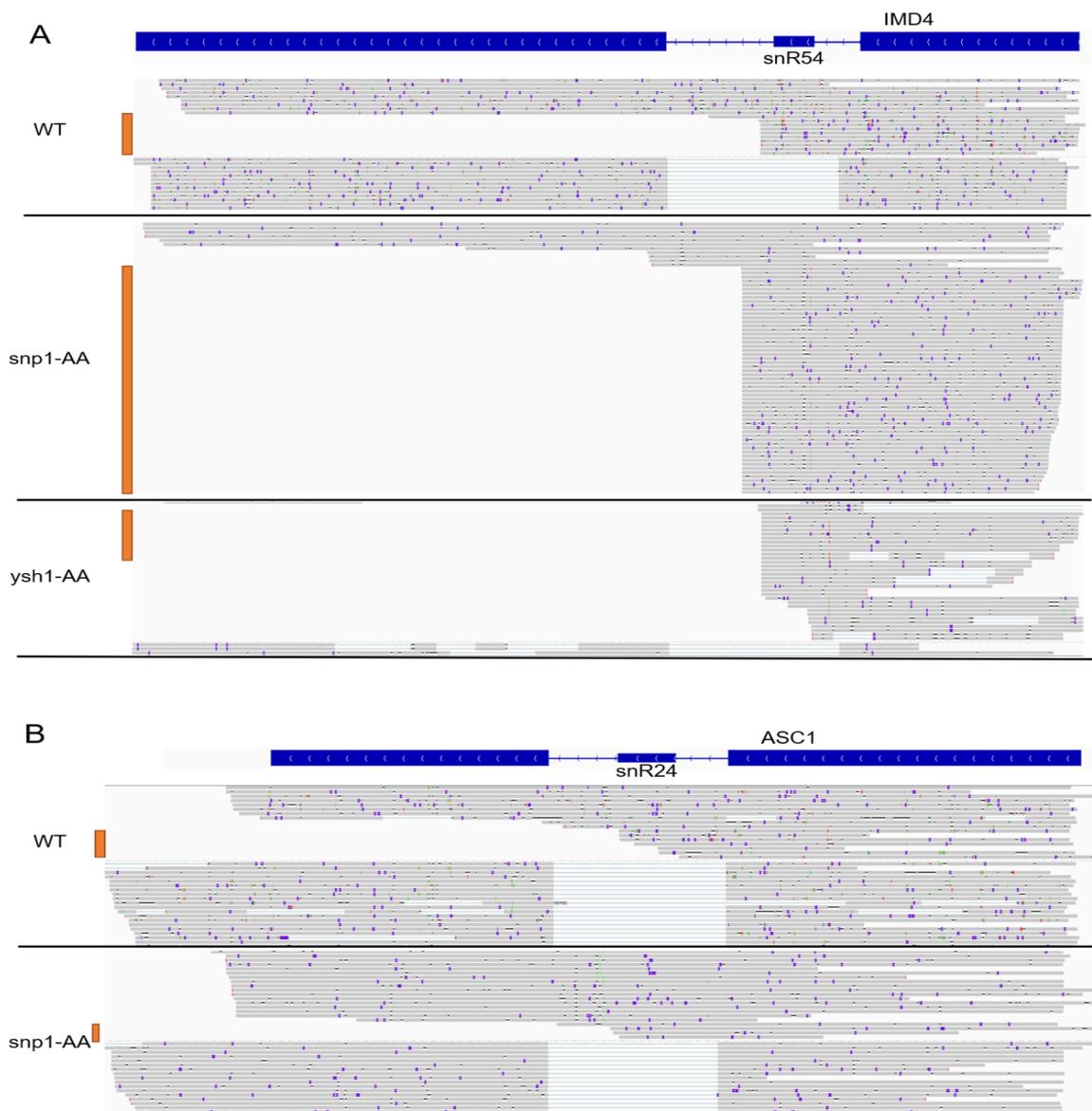
In order to study the genome-wide effects of defects in a range of RNA processing pathways, we performed both direct RNA NanoPore sequencing of wild-type, CPA deficient, splicing deficient, as well as cDNA NanoPore sequencing (not shown) of fast and slow RNA Pol II mutant strains of *S. cerevisiae*. These NanoPore sequencing data are poly(A)-tail based, meaning that any RNA sequenced must have a poly(A) tail with demonstrated detection down to at least 10As in length. Surprisingly, in every sample studied, we detect the presence of multiple hmsnoRNA reads (Figure 5.XA), including the aforementioned ASC1/snR24 hmsnoRNA in some samples (Figure 5.XB) without prior in vitro polyadenylation. This in contrast to our previous work in which we detected hmsnoRNA by Nanopore sequencing upon inactivation of the splicing factor Slu7 after in vitro polyadenylation.

These data apparently contradict our previous finding that these hmsnoRNAs were not polyadenylated. This contradiction could have resulted for a couple of reasons. Firstly, these polyadenylated hmsnoRNA species are detected by NanoPore at low levels, such as the ASC1/snR24 hmsnoRNA only having a handful of reads. It is possible that for these hmsnoRNAs the vast majority of the species are not polyadenylated and that the particularly low count of poly(A) hmsnoRNAs were present in the Northern Blot, but not detectable. Alternatively, the failure to detect these poly(A)<sup>+</sup> hmsnoRNAs could be a technical limitation of the technique used to separate the poly(A)<sup>+</sup> and poly(A)<sup>-</sup> fractions prior to Northern Blot

analysis. Oligo-d(T)<sub>25</sub> beads are primarily designed for the purification of mRNAs, which generally carry long poly(A) tails added by the poly(A) polymerase of their respective CPA complex. However, snoRNA 3' ends are typically polyadenylated by the TRAMP complex to assist in targeting the snoRNA for 3' end trimming by the nuclear exosome. These poly(A) tails added by the TRAMP complex are generally much shorter than their PAP dependent contemporaries: often only being single digits into the low teens of adenosines long. This difference between the designed poly(A) tail length target of oligod(T)<sub>25</sub> beads and the expected TRAMP-dependent poly(A) tail lengths could result in the failure for hmsnoRNAs to bind to the beads during purification, which could explain their exclusive presence in the poly(A)- fraction of the previous experiment. NanoPore sequencing technology, however, is known to be capable of sequencing short poly(A) tails as previously mentioned. This potential increased sensitivity for shorter poly(A) tails in comparison to oligo d(T)<sub>25</sub> would then explain why these poly(A)+ hmsnoRNAs are detected via NanoPore sequencing, but not in the previous Northern Blot. Additionally, it is possible that each hmsnoRNA species may have differing levels of polyadenylation and/or stability, meaning our previous species specific experimentation could have coincidentally targeted hmsnoRNAs that happened to not be detectable easily by poly(A)+ Northern Blot, while other hmsnoRNAs could have.

Regardless of the reasoning as to why these hmsnoRNA's poly(A) status was previously missed, these data provide both more questions, and potentially a better way to study these RNAs than our previous methodology. First-most, in order to determine both if hmsnoRNAs are generally polyadenylated or if there are varying ratios of poly(A) populations for each hmsnoRNA, I propose a paired set NanoPore sequencing experiments. In these two sequencing experiments we would split one total RNA sample into two parts; one which would be sequenced without further treatment with direct RNA NanoPore sequencing, the other we would process as we had previously i.e. ribodepletion followed by *in vitro* polyadenylation prior

to direct RNA NanoPore sequencing. We would then be able to compare the presence of every hmsnoRNA between the two samples, which should indicate the ratio of poly(A)+ and poly(A)- hmsnoRNAs for each hmsnoRNA.



**Figure 5.5. hmsnoRNAs are detectable by NanoPore sequencing. A.** Direct RNA NanoPore sequencing data for wild-type-AA, *snp1-AA*, and *ysh1-AA* at the *IMD4* locus. Orange boxes indicate reads that represent the hmsnoRNA *IMD4/snR54*. cDNA NanoPore sequencing of wild-type, Fast Pol II, and slow Pol II are not included, but also show detectable levels of the *IMD4/snR54* hmsnoRNA. **B.** Direct RNA NanoPore sequencing of wild-type-AA, and *snp1-AA* of the *ASC1* locus. Orange boxes indicate reads that correspond to the *ASC1/snR24* hmsnoRNA