THE ROLE OF RTR1 AND RRP6 IN RNAPII TRANSCRIPTION TERMINATION

Melanie J. Fox

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Biochemistry and Molecular Biology
Indiana University

October 2015

Accepted by the Graduate Faculty, of Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Amber L. Mosley, Ph.D., Chair

Mark G. Goebl, Ph.D.

Doctoral Committee

Yunlong Liu, Ph.D.

August 31, 2015

Ronald C. Wek, Ph.D.

DEDICATION

To my husband, Justin King, who woke me with coffee in the morning and welcomed me home with a beer at night.

To the greatest dog in the world, Rupert Fox. My sidekick. My best friend.

ACKNOWLEDGEMENTS

This thesis was only possible through the guidance and support and many individuals. I would like to express my sincerest gratitude to:

My mentor, Dr. Amber Mosley, for her guidance, support, and friendship throughout this work. She always encouraged me to think outside the box and follow my heart, in both research and life.

The members of my committee, Dr. Mark Goebl, Dr. Yunlong Liu, and Dr. Ron Wek, for their invaluable advice on my project.

Everyone in the Mosley Lab for support over the years. Working with fun, helpful people made the difficult times bearable and the good times a blast. Whitney Smith-Kinnaman, Jerry Hunter, Megan Zimmerly, Jason True, Mary Cox, Michael Berna, Gigi Cabello, Lynn Bedard, Rachel Chan, Elizabeth DeVlieger, Asha Boyd, Gabi Mazur, Sarah Peck, and Jose Victorino.

Many members of my family and close friends I consider family who have supported me endlessly during the best and worst of times.

Everyone in the Department of Biochemistry and Molecular Biology, including Dr. Charlie Dong, Dr. Nuria Morral, Dr. Maureen Harrington, Dr. Tim Corson, Dr. Mark Goebl, Jack Arthur, Sandy McClain, Melissa Tarrh, Sheila Reynolds, and Darlene Lambert for equipment use and support.

The founding members of Central Indiana Science Outreach and the Staff and Volunteers in the Dinosphere at the Children's Museum of Indianapolis for always reminding me that science is a lot of fun.

Melanie J. Fox

THE ROLE OF RTR1 AND RRP6 IN RNAPII TRANSCRIPTION TERMINATION

RNA Polymerase II (RNAPII) is responsible for transcription of messenger RNA (mRNA) and many small non-coding RNAs. Progression through the RNAPII transcription cycle is orchestrated by combinatorial posttranslational modifications of the C-terminal domain (CTD) of the largest subunit of RNAPII, Rpb1, consisting of the repetitive sequence $(Y_1S_2P_3T_4S_5P_6S_7)_n$. Disruptions of proteins that control CTD phosphorylation, including the phosphatase Rtr1, cause defects in gene expression and transcription termination. There are two described RNAPII termination mechanisms. Most mRNAs are terminated by the polyadenylation-dependent cleavage and polyadenylation complex. Most short noncoding RNAs are terminated by the Nrd1 complex. Nrd1-dependent termination is coupled to RNA 3' end processing and/or degradation by Rrp6, a nuclear specific subunit of the exosome. The Rrp6-containing form a 3'-5' exonuclease complex that regulates diverse aspects of nuclear RNA biology including 3' end processing and degradation of a variety of noncoding RNAs (ncRNAs). It remains unclear whether Rrp6 is directly involved in termination. We discovered that deletion of *RRP6* promotes extension of multiple Nrd1-dependent transcripts resulting from improperly processed 3' RNA ends and faulty transcript termination at specific target genes. Defects in RNAPII termination cause transcriptome-wide changes in mRNA expression through transcription interference and/or antisense repression, similar to previously reported effects of Nrd1 depletion from the nucleus. Our data indicate Rrp6 acts with Nrd1 globally

to promote transcription termination in addition to RNA processing and/or degradation. Furthermore, we found that deletion of the CTD phosphatase Rtr1 shortens the distance of transcription before Nrd1-dependent termination of specific regulatory antisense transcripts (ASTs), increases Nrd1 occupancy at these sites, and increases the interaction between Nrd1 and RNAPII. The *RTR1/RRP6* double deletion phenocopies an *RRP6* deletion, indicating that the regulation of ASTs by Rtr1 requires Rrp6 activity and the Nrd1 termination pathway.

Amber L. Mosley, Ph.D., Chair

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

4tU-seq	4-thiouracil labeling with RNA sequencing
ALS4	Amyotrophic lateral sclerosis type 4
AOA2	Ataxia oculomotor apraxia type 2
APA	Alternative polyadenylation
AS	Antisense
ATP-^{32}P	Adenosine triphosphate labeled with radioactive ^{32}P
bp	Base pair
CBC	Cap binding complex
CF	Cleavage factor
ChIP	Chromatin immunoprecipitation
ChIP-chip	Chromatin immunoprecipitation on microarray
ChIP-exo	Chromatin immunoprecipitation with exonuclease digestion
ChIP-seq	Chromatin immunoprecipitation with DNA sequencing
CID	CTD interacting domain
CML	Chronic myelogenous leukemia
CPF	Cleavage and polyadenylation factor
CRAC	Crosslinking and analysis of cDNAs
CTD	C-terminal domain
CTDK1	CTD kinase 1
CUTs	Cryptic unstable transcripts

DECOID	Decreased expression of complexes by overexpression of interacting domains
DNA	deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
ET	Extended transcript
FDR	False discovery rate
GO	Gene Ontology
lncRNA	Long noncoding RNA
mRNA	messenger RNA
MUTs	Meiotic unstable transcripts
NEXT	Nuclear exosome targeting complex
NNS	Nrd1-Nab3-Sen1 termination machinery
nt	Nucleotide
NUTs	Nrd1-unterminated transcripts
OD_{600}	Optical density measured at a wavelength of 600 nm
ORF-T	Open reading frame transcript
P#	Preparation number
PAR-CLIP	Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation
PCH1	Pontocerebellar hypoplasia type 1
PCR	Polymerase chain reaction
PNK	Polynucleotide Kinase
RIN	RNA integrity number

RNA	ribonucleic acid
RNAPII	RNA Polymerase II
RRM	RNA recognition motif
rRNA	ribosomal RNA
SDS	Sodium dodecyl sulfate
Ser2-P	Serine 2 phosphorylation
Ser5-P	Serine 5 phosphorylation
Ser7-P	Serine 7 phosphorylation
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SUTs	Stable unannotated transcripts
TFIIH	Transcription factor II H
TOV	Terminator override
ts	Temperature sensitive
TSS	Transcription start site
TTS	Transcription termination site
Tyr1-P	Tyrosine 1 phosphorylation
USE	Upstream Sequence Element
UTR	Untranslated region
WT	Wild type
XUTs	Xrn1-unterminated transcripts
YPD	Yeast extract-peptone-dextrose medium

INTRODUCTION

1. Transcription by RNAPII and Regulation of the CTD

RNA Polymerase II (RNAPII) is responsible for the transcription of messenger RNA (mRNA) from protein coding genes and several forms of small non-coding regulatory RNAs. RNAPII transcription is divided into three steps, initiation wherein RNAPII, aided by general and gene-specific transcription factors, is recruited to the promoter of genes, facilitating initial *de novo* synthesis of the nascent pre-mRNA. Following initiation, RNAPII proceeds to elongation where the bulk of the RNA synthesis occurs, following by termination of transcription. A scheme of terms used to describe transcribed regions throughout this thesis is shown in Figure 1.

This thesis focuses on key events regulating transcription elongation and termination, and how these events modulate the transcriptome. Recruitment of factors involved in transcription elongation and termination is orchestrated by the phosphorylation state of the C-terminal Domain (CTD) of the largest subunit of RNAPII, Rpb1, which has the repetitive sequence (Tyrosine^{1}-Serine^{2}-Proline^{3}-Threonine^{4}-Serine^{5}-Proline^{6}-Serine^{7})$_{n}$ where n = 26 in yeast and n = 52 in humans (Reviewed in [1]). The phosphorylation pattern of the CTD sequence has been shown to be specific to the location of RNAPII along the transcribed region of the gene (Figure 2) [2]. For example, in the model eukaryotic organism *Saccharomyces cerevisiae* it has been shown that the RNAPII CTD is hypophosphorylated when the initiation complex forms at the promoter. Once

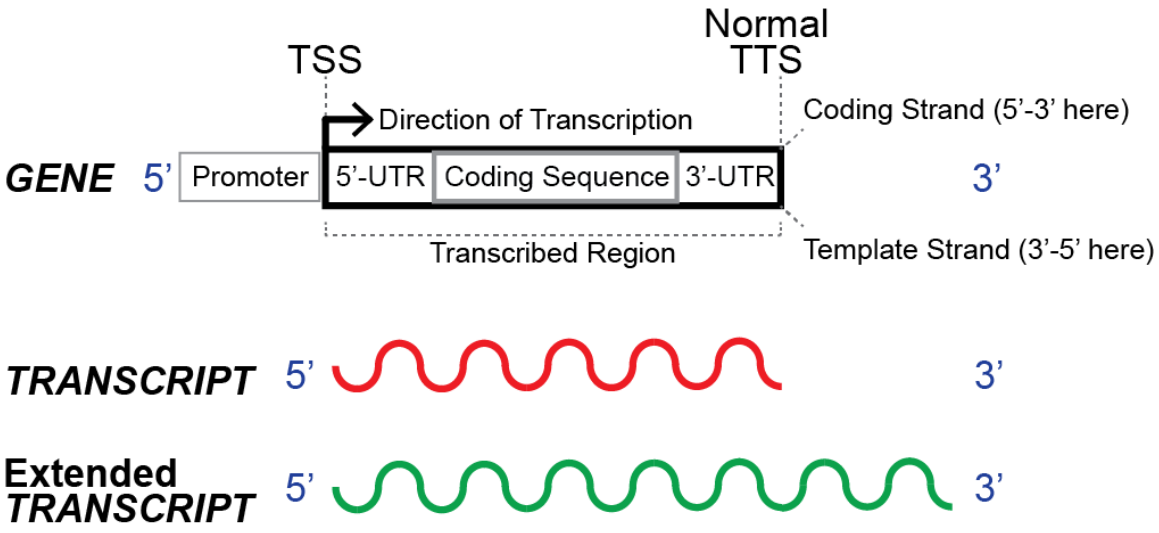

Figure 1: Schematic of terms. This thesis focuses on the transcribed region of genes, from transcription start site (TSS) to transcription termination site (TTS), and the resulting RNA transcripts. When referring to the 5'-end or 3'-end of a gene, this is regarding the orientation of the coding strand and the RNA, as shown here. An extended transcript, as discussed here, is assumed to start at the same TSS and have a longer 3'-end due to improper termination. Gene and transcript names are always written in italics. Protein names are capitalized and not italicized.

elongation begins, serine 5 phosphorylation (Ser5-P, in Figure 2) is highest at the beginning of the transcription. However, during the early phase of transcription elongation, levels of Ser5-P decrease as RNAPII travels towards the transcription termination site [3]. Coincident with lowered Ser5-P during transcription, Ser2 of the CTD of RNAPII becomes phosphorylated so that by the mid- to late- phase of transcription elongation there is a mixture of both Ser5- and Ser2-phosphorylation. Ultimately, Ser2-P RNAPII predominates at the end of transcription of long genes (generally over 1000bp) (Figure 2) [4]. Ser7 has also recently been shown to be phosphorylated, peaking at the beginning of the gene similarly to Ser5-P with a slight increase at the end of transcription with Ser2-P. Together these phosphorylation patterns in the CTD of RNAPII serve as a binding platform to recruit the protein complexes required for each stage of transcription, facilitating association between the transcription machinery and key protein complexes that carry out mRNA capping, pre-mRNA splicing, cleavage of the nascent RNA, and polyadenylation [2].

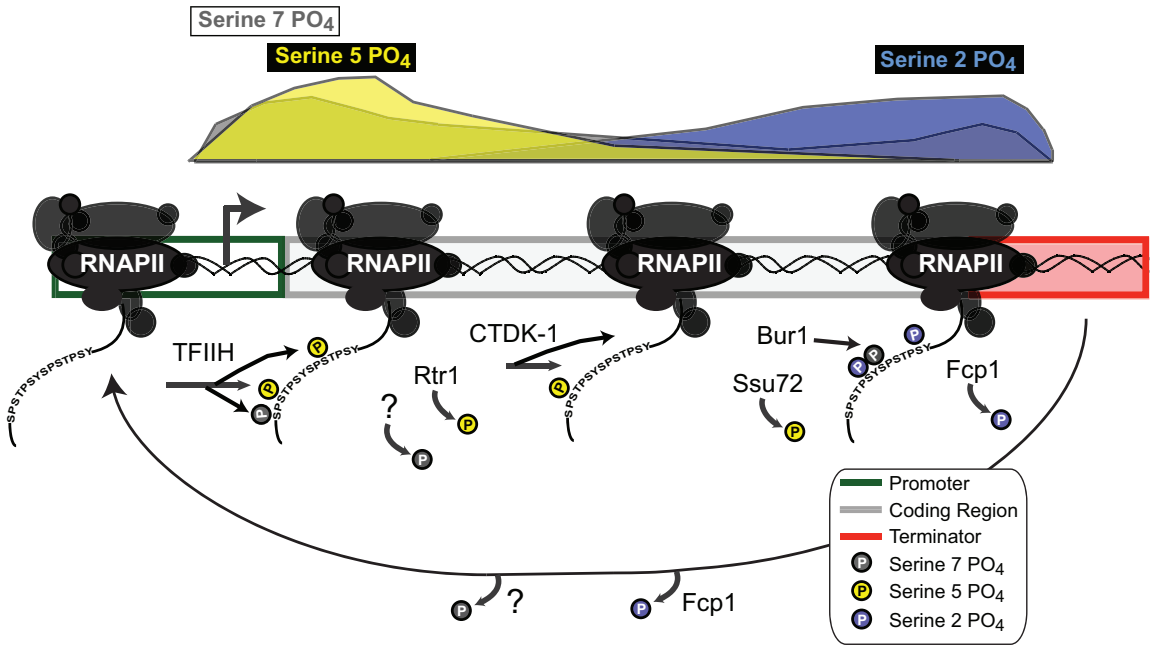

Figure 2: Schematic of RNAPII CTD phosphorylation patterns across an average gene in *S. cerevisiae*. The phosphorylation state of the CTD (each site is represented by a different color as indicated in the legend on the bottom right) as well as the major known kinases and phosphatases are shown.

Collectively, protein kinases and phosphatases that add and remove Ser2, Ser5, and Ser7 phosphorylation marks in the CTD of RNAPII determine the levels of CTD modification and efficiency of RNAPII transcription. As illustrated in Figure 2, Ser5-P and Ser7-P are added early in elongation by TFIIH, specifically its Kin28 subunit [5-7]. In the mid- to late phase of transcription elongation, some Ser5-P is removed by Rtr1, coordinate with Ser2-P addition by Ctk1, a subunit in the CTDK1 complex [8-10]. The protein kinase Bur1 adds new Ser7-P as RNAPII nears the end of gene transcription [11]. Subsequently, the protein phosphatase Ssu72 removes Ser5-P and Ser7-P, and Fcp1 eliminates Ser2-P [11] in a step that is proposed to be important for recycling RNAPII for initiation of another round of transcription. It is not clear why three separate protein phosphatases are required for regulation of RNAPII CTD when there is apparent overlap in their functions. There is also uncertainty about how the protein phosphatases interact physically and/or genetically to orchestrate their activities on transcription elongation.

How do these changing compositions of CTD phosphorylation affect transcription elongation and termination? Are there CTD modifications yet to be discovered that affect these key transcription processes? Two post-translational modifications of proteins have recently been proposed to play important roles in termination of RNAPII transcription. As noted above, the regulation of the CTD phosphorylation state requires interplay between protein kinases and phosphatases, but has also been suggested to involve proline isomerases [2]. The Ser5 phosphatase Ssu72 binds the CTD of RNAPII when Pro6 of the CTD is

in *cis* conformation following recruitment of the proline isomerase Ess1 [12]. Second, *in vivo* studies have indicated that Tyr1-P of the CTD may modify the way that termination factors interact with Ser2-P and RNAPII [13]. Using ChIP-chip, Ser2-P and Tyr1-P were seen to increase as RNAPII elongation proceeds across the length of genes based studies featuring on genome-wide averaging of 997 genes classified by length. Tyr1-P signals were found to peak *in vivo* about 180 nucleotides (nt) upstream of the poly(A) site and then were significantly lowered, whereas Ser2-P occurred at the same point during transcription elongation but were maintained throughout- 200 nt downstream of the poly(A) site [13]. Although Ser2-P levels were consistent across this region, CTD-interacting proteins were not recruited until Tyr1-P signals dropped, leading the authors to hypothesize that Tyr1-P of RNAPII inhibits recruitment of the termination factors to “fine-tune” the timing of transcription termination [13].

Phosphorylation at Tyr1 may be removed by the phosphatase Glc7 in yeast [14]. Glc7 is a type 1 serine/threonine protein phosphatase that is also a member of the cleavage and polyadenylation factor complex along with the RNAPII CTD phosphatase Ssu72 [15, 16]. Of interest, it has also been shown that Tyr1-P can be dephosphorylated *in vitro* by the atypical phosphatase Rtr1 [17]. Additional studies are needed to determine the role of these protein phosphatases in the regulation of Tyr1-P levels and overall RNAPII transcription *in vivo*. In summary, multiple protein modifications of the CTD of RNAPII, including protein phosphorylation at serine and tyrosine and proline

isomerization, are suggested to modulate elongation and termination phases of transcription.

2. RNAPII Transcription Termination

In the model organism, *Saccharomyces cerevisiae*, two mechanisms are suggested to terminate RNAPII transcription. These mechanisms involve the Cleavage and Polyadenylation Factor (CPF) complex or Nrd1, along with specific sequences embedded within the nascent RNA (Figure 3) [4, 18-20]. Table 1 highlights a list of protein complexes and their subunits involved in RNAPII transcription termination. Current ideas about the functions of these proteins and RNA sequences directing transcription termination will be discussed more fully below.

Yeast Complex	Yeast Protein	Human Homolog	Protein Function	Role in P(A) Dependent Termination	Role in Alternate Termination	Ref
CPF	Cft1	CPSF160	Binds RNAPII Ser5-P and Ser2-P CTD, Binds Poly(A) site RNA	Promotes RNAPII pausing and RNA cleavage for Rat1 entry	N.D.	21
	Cft2	CPSF100	Interacts with RNAPII CTD, Binds cleavage site RNA, Bridges CPF with Pcf11	Promotes RNAPII pausing and RNA cleavage for Rat1 entry	N.D.	22
	Fip1	FIP1	Bridges Pap1 with rest of CPF	Interacts directly with Pap1 to improve polyadenylation	N.D.	23
	Glc7	PP1	Ser-Thr phosphatase, Dephosphorylates Sen1	N.D.	Promotes Sen1 Recruitment and/or helicase activity	21
	Mpe1	-	RNA-binding protein	Promotes link between CPF complex and pre-mRNA	N.D.	24
	Pap1	PAP1	RNA polyadenylation, implicated in RNA export	Polymerase responsible for polyadenylation of mRNA	N.D.	25
	Pfs2	-	Scaffolding protein	Promotes formation of CPF	N.D.	26
	Pta1	Symplekin	Scaffolding protein, Bridges CF1A and CPF	N.D.	Maintains integrity of CPF	21
	Pti1	-	Scaffolding protein, Bridges Ref2 and Pta1	N.D.	Maintains integrity of CPF	27
	Ref2	-	RNA-binding protein, Scaffolding Protein, Bridges Pti1 and Glc7	N.D.	Maintains integrity of CPF	21

Yeast Complex	Yeast Protein	Human Homolog	Protein Function	Role in P(A) Dependent Termination	Role in Alternate Termination	Ref
CPF	Ssu72	SSU72	RNAPII Ser5-P CTD Phosphatase	Promotes recruitment of Pcf11 to RNAPII	Promotes recruitment of Pcf11 to RNAPII	21
	Swd2	WDR82	Scaffolding protein, Subunit of histone methylation complex Set1	N.D.	Maintains integrity of CPF	21
	Syc1		Scaffolding protein	Maintains integrity of CPF	Maintains integrity of CPF	21
	Ysh1	CPSF73	Endoribonuclease, Cleaves poly(A) site RNA	Provides entry point for Rat1	Provides entry point for exoribonuclease	21
	Yth1	CPSF30	Binds poly(A) site RNA and RNAPII	Promotes RNAPII pausing	N.D.	21
NNS	Nrd1	SCAF8 and SCAF4	Binds RNAPII Ser5-P CTD, RNA-binding Protein	No effect observed	Bridges CTD and RNA to release RNAPII, Recruits Sen1 to RNAPII	21
	Nab3	-	RNA-binding protein	N.D.	Bridges Nrd1 and Sen1	21
	Sen1	Senataxin	Binds RNAPII CTD, 5'-3' RNA:DNA helicase	Promotes Rat1 activity by exposing RNA	Unwinds RNA:DNA hybrid in RNAPII	21
	Sto1 (aka Cbc1)	CBP80	Member of Cap-binding protein, binding partner of Cbc2	Acts with Npl3p to export mRNA	Binds AT to promote degradation of downstream RNA	28
	Cbc2	CBP20	Member of Cap-binding protein, binding partner of Sto1	Acts with Npl3p to export mRNA	Binds AT to promote degradation of downstream RNA	28
CF1a	Pcf11	PCF11	Binds RNAPII Ser2-P CTD, Scaffolding protein	Promotes RNA cleavage and Rat1 recruitment, Bridges CTD and RNA to release RNAPII	Bridges CTD and RNA to release RNAPII	21

Yeast Complex	Yeast Protein	Human Homolog	Protein Function	Role in P(A) Dependent Termination	Role in Alternate Termination	Ref
CF1a	Rna14	CSTF77	Scaffolding protein	Promotes RNA cleavage	Maintains integrity of CF1A	21
	Rna15	CSTF64	Binds Poly(A) site RNA	Promotes RNA cleavage and Rat1 recruitment	Maintains integrity of CF1A	21
	Clp1	CLP1	Scaffolding protein, Bridges CF1A and CPF	Contributes to CF1A formation and interaction with CPF	N.D.	29
	Hbs1 (?)	HBS1L	GTPase, implicated in mRNA no-go decay			29
	Yra1 (?)	-	RNA binding protein, involved in RNA export	Couples mRNA export to 3'processing through interaction with Pcf11	N.D.	29
	Hrp1 (?)	-	RNA binding protein	Binds poly(A) sequence	Maintains integrity of CF1A	29
Exo-nuclease	Rtt103	-	Binds RNAPII Ser2-P CTD, Bridges Rat1 to RNAPII CTD	Recruits Rat1 and Pcf11 to RNAPII	N.D.	21
	Rat1	XRN2	5'-3' Exoribonuclease, Degrades Ysh1-generated downstream cleavage product	Promotes Pcf11 and Rna15 recruitment, Collides with RNAPII near RNA exit channel	No effect observed	21
	Rai1	DOM3Z	De-capping endoribonuclease, Pyrophospho-hydrolase	Promotes Rat1 stability and activity	N.D.	21

Table 1: Relevant proteins known to be involved in RNAPII transcription termination. Table adapted from [21] with updated information from Saccharomyces Genome Database (http://www.yeastgenome.org) and listed references.

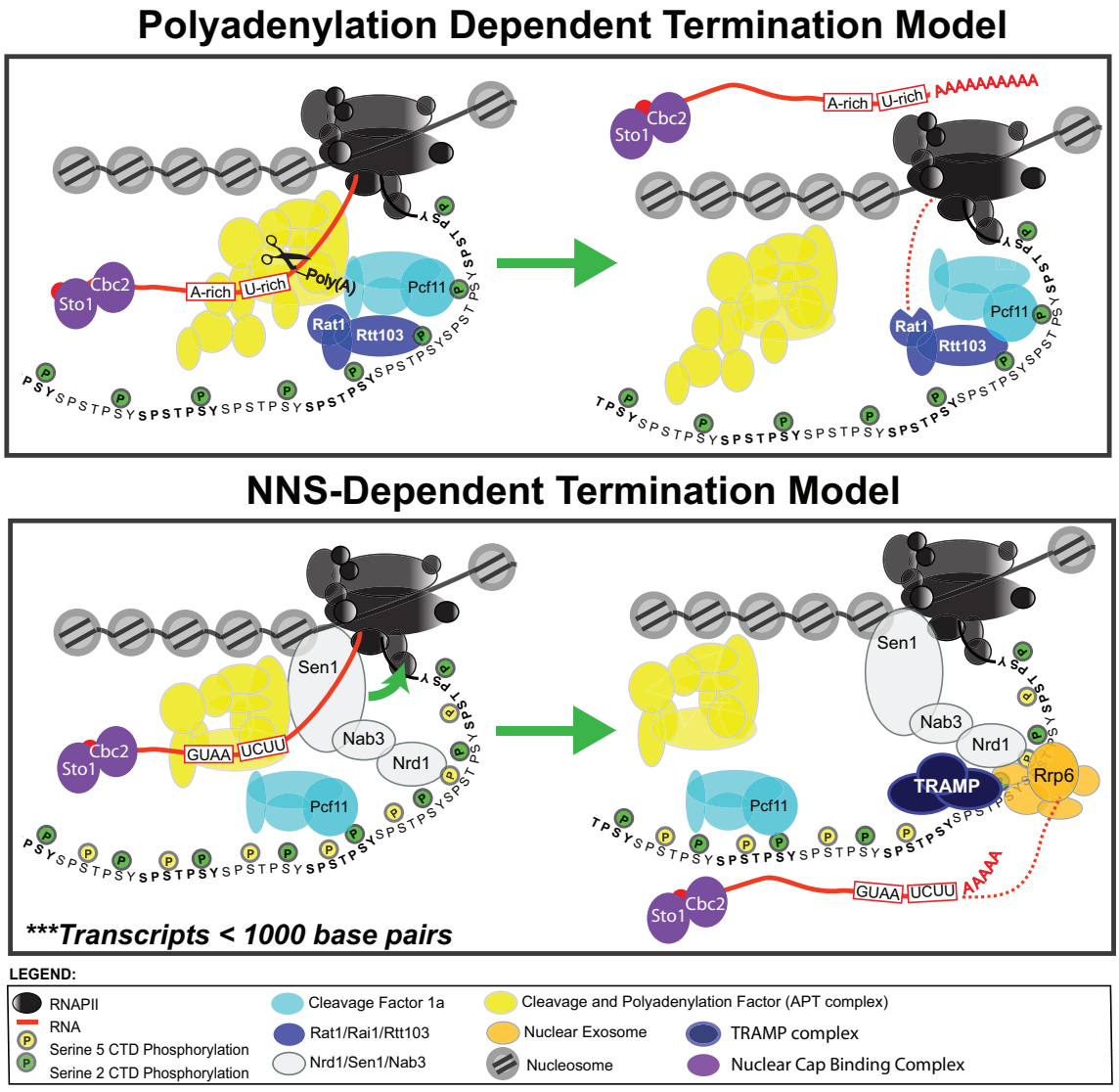

Figure 3: Representation of the two current models of RNAPII transcription termination. Top panel: Polyadenylation-dependent model. Bottom panel: Nrd1-Nab3-Sen1 (NNS)-dependent termination.

2.1. Cleavage and Polyadenylation Dependent Termination

CPF in complex with Cleavage Factor Ia (CFIa) is thought to facilitate general termination of long (>1000 bp) polyadenylated mRNAs by a mechanism integrating the polyadenylation process [15, 22, 30-32] [33]. Current models of RNAPII transcription termination suggest a coupling of transcription of the 3' end of the transcript to the recruitment of 3' end processing proteins such as the CPF. RNAPII pausing due to transcription of the poly(A) site facilitates recruitment of 3' end processing machinery [34, 35]. It has been proposed upon transcription of the 3'-portion of genes that Ser2-P on the CTD of RNAPII recruits Rtt103 through a CTD-Interacting Domain (CID). Recruitment of Rtt103 to the transcribing RNAPII is suggested to cooperate with the CID of the protein Pcf11, which then culminates to recruit the CFIa complex [32]. An additional feature of this model is that Tyr1-P of the RNAPII CTD may inhibit recruitment of the termination complexes. In this view, the possible dephosphorylation of Tyr1 by the Glc7 phosphatase may help to fine-tune timing of transcription termination [14]. Such regulation could facilitate processes such as alternative polyadenylation site usage, a known regulator of mRNA transcript stability [36]. At the same time, CPF associates with paused RNAPII and recognizes the poly(A) signal sequence [37].

Recruitment of Rtt103 is thought to also facilitate association of additional proteins, including the 5' → 3' exoribonuclease Rat1 and its activating partner Rai1. Rat1 has been proposed to degrade the nascent RNA downstream of the CPF cleavage site while the RNA is still attached to RNAPII, causing the

disassociation of RNAPII, an event termed the “torpedo termination model” [38, 39].

2.2. Alternative Termination through the NNS pathway

One way to address the specific roles of Rtr1 and Ssu72 is to determine the effects of loss of Rtr1 activity on pathways dependent on proper regulation of Ser5-P. For example, we previously showed that deletion of *RTR1* results in increased Ser5-P toward the transcription termination site, leading to disruption of termination at studied genes [8]. As highlighted above, transcription termination is carried out by two pathways orchestrated by the CTD code. Transcription of long, polyadenylated mRNAs are terminated in a cleavage and polyadenylation-dependent mechanism that occurred independent of Rtr1 and Ssu72 [15, 22, 30-32]. Shorter RNAPII transcripts (<1000 bp) including some small protein-coding genes, snRNAs, and many types of non-coding RNAs are terminated by an alternative Nrd1, Nab3, and Sen1 (NNS) -dependent mechanism [20, 40-44].

In the current model of NNS-dependent termination, Nrd1 and Nab3 are recruited to the termination complex by the CID of Nrd1 that binds the Ser5-P CTD, and RNA Recognition Motifs (RRM Domains) present in Nrd1 and Nab3 that recognize specific sequences in the nascent RNA [4, 19, 45-50]. The helicase Sen1 participates by unwinding RNA:DNA hybrids formed by the nascent RNA and template strand promoting RNAPII termination [51-54]. RNAPII transcribes beyond the 3’ end of the functional RNA into the “termination zone” of

the gene before transcription is actually terminated [51]. It has been shown *in vitro* that an interaction between Sen1 and RNAPII alone can lead to transcription termination by facilitating RNAPII release from the DNA template and synthesizing RNA, suggesting that Sen1 can trigger termination independently of Nrd1 and Nab3 [51, 55]. These data suggest that *in vivo* Nrd1 and Nab3 may aid in the recruitment of Sen1 to termination sites of genes or expedite the process, but Nrd1 and Nab3 are not required *per se* to carry out the termination reaction.

The nascent terminated RNA is then polyadenylated by the TRAMP complex, and the nuclear exosome trims the ends of stable transcripts, such as snRNAs and SUTs (Stable Unannotated Transcripts) or completely degrades the transcript in the case of CUTs (Cryptic Unstable Transcripts) (Figure 3) [20, 42, 43, 51, 56-60]. The Nrd1 CID has also been shown to mediate the interaction between the NNS complex and the TRAMP subunit Trf4, coordinating transcription termination with RNA degradation and processing [61]. The Nrd1-Nab3 complex also requires the function of the CFIa subunit Pcf11. Disruption of the Ser2 binding CID of Pcf11 leads to Nrd1 retention on chromatin suggesting that wild-type Pcf11 prevents the localization of Nrd1 to terminator distal (3') regions [62]. Competition for CTD-binding between Nrd1 and Pcf11 may be regulated by the CTD prolyl isomerase Ess1 [63]. Ess1 isomerizes the S5-P-P6 bond resulting in a conformational change favored by Ssu72. Subsequent removal of S5-P by Ssu72 releases Nrd1 from the CTD and may allow Pcf11 to bind adjacent S2-P. It has been suggested that this ordered Nrd1-binding, then

release, followed by Pcf11 recruitment is somehow required for NNS-dependent termination [63]

3. The Nuclear RNA Exosome

The RNA exosome is a 3'-5' ribonuclease complex involved in termination of short RNAs, RNA quality control surveillance, and RNA degradation (first described in [64]). The core of the RNA exosome is a multi-subunit complex similar in structure to the proteasome that features a channel that leads the RNA substrate to the catalytic subunit Dis3 (also known as Rrp44), similar in structure to the proteasome [65-68]. In yeast, there are two forms of the RNA exosome: the cytoplasmic exosome consists of the core complex only, and the nuclear exosome contains of the core exosome with the addition of Rrp6. The human RNA exosome exists in three forms localized to the cytoplasm, nucleoplasm, and nucleolus, with the Rrp6-containing complex localized to the nucleolus (reviewed in [69]).

Although Rrp6 is not essential for viability in yeast, it has been shown to be required for proper 3' end trimming of primary sn/snoRNAs [45, 70], degradation of short-lived, non-coding CUTs [71-75] and improperly terminated RNAs [70, 76], as well as regulation of polyA tail length [77, 78] and termination of specific short transcripts. Recent studies using high-resolution tiling arrays found that Dis3 and Rrp6 have both shared and distinct roles in the degradation and processing of various RNAs [79, 80]. Rrp6 has also been implicated in the termination and 3' end processing of a variety of noncoding RNAs, most that are

not terminated by traditional polyA-dependent termination mechanisms [43, 74]. Furthermore, a number of studies have uncovered important mechanistic details regarding the cellular implications of transcriptome regulation by Rrp6 and the RNA exosome. Rrp6 and other exosome subunits are required for proper meiosis [81] and cellular differentiation [82]. Additionally, inhibition of exosomal proteins, either by mutation or production of autoantigens, can lead to a variety of human diseases [83-86]. Here we will discuss the latest research on the structure and function of the RNA exosome and key associating subunits, such as Rrp6 and Dis3.

3.1. Nuclear Exosome Structure in Yeast

Exo9, the Core Barrel

The nine-subunit non-catalytic core is organized in to a central channel that directs the RNA to the Dis3 nuclease. As noted above, the RNA exosome has a barrel structure similar to the proteasome with a core that is composed of two stacked rings lacking nuclease activity [87]. The bottom ring includes six proteins with homology to RNase PH, a phosphorylitic 3'-5' exonuclease, but amino acid changes in the active sites in eukaryotes render them inactive. These RNase PH-like proteins include Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr3. The top ring includes cap proteins containing S1 and KH domains that are often found in RNA binding proteins [88-90]. These ring proteins form a stable core that adjoin with cap proteins Rrp4, Rrp40, and Csl4. The RNase PH-like proteins form a stable core with the cap proteins Rrp4 and Rrp40, and a third cap protein,

Csl4, which is more loosely bound [91-93]. These proteins are well conserved among eukaryotes, and Table 2 lists the names of the homologous subunits from yeast and humans.

Dis3, aka Rrp44

In the cytoplasm of *Saccharomyces cerevisiae,* the core Exo9 barrel is associated with the Dis3 exonuclease ($Exo10^{Dis3}$). All Exo10 subunits (barrel plus Dis3) are essential in yeast [64, 107-109]. Dis3 is the only catalytic subunit of the RNA exosome that is a part of the exosome complex in both the cytoplasm and nucleus. It binds with the RNase PH proteins in the Exo9 core structure on the opposite side of the barrel from the cap proteins [90, 92]. It has processive exoribonuclease activity through RNase II-like domains as well as endoribonuclease activity in its PIN domain [87, 94, 110-112]. An *in vivo* RNA crosslinking study using cleavable proteins (described below) has shown that Dis3 is capable of using both endonuclease and exonuclease activities on most substrates [80]. The authors used the CRAC (crosslinking and analysis of cDNAs) technique to purify protein-RNA complexes. By utilizing cleavage sites introduced in the target proteins, these “split-CRAC” studies identified specific domains in the exosome subunits that interact with the RNA substrates to determine their individual contributions to RNA decay. After crosslinking the full-length protein to its RNA targets, Dis3 was cleaved into two fragments, separating the PIN and RNB domains. The transcriptome-wide profiles of RNA substrates bound to these fragments were compared to substrates bound to full-

length Dis3. All three profiles were found to be similar, indicating that the same

Complex			*S. cerevisiae*	Human	Description	Disease Association
Exo9	Exo10	Exo11	Rrp41	Rrp41 (EXOSC4, Ski6)	RNase PH-like protein in barrel [90]	-
			Rrp42	Rrp42 (EXOSC7, EAP1)	RNase PH-like protein in barrel [90]	-
			Rrp43	Rrp43 (OIP2, EXOSC8)	RNase PH-like protein in barrel [90]	Pontocerebellar hypoplasia [85]
			Rrp45	Rrp45 (PM/Scl75, EXOSC9)	RNase PH-like protein in barrel [90]	Autoantigen in polymyositis, scleroderma, and PM/Scl overlab disease [98-100]
			Rrp46	Rrp46 (EXOSC5)	RNase PH-like protein in barrel [90]	Up-regulated in lung cancer, melanoma, and prostate cancer [101]
			Mtr3	Mtr3 (EXOSC6)	RNase PH-like protein in barrel [90]	-
			Rrp4	Rrp4 (EXOSC2)	S1/KH cap proteins [90]	-
			Rrp40	Rrp40 (EXOSC3)	S1/KH cap proteins [90]	Mutations in pontocerebellar hypoplasia [84,102]
			Csl4	Csl4 (EXOSC1)	S1/KH cap proteins [90]	-
			Dis3 (Rrp44)	DIS3 (EXOSC11)	3'-5' Endo/Exonuclease, localized in the nucleus [80]	Mutations in multiple myeloma [43,44]
				DIS3L1	3'-5' Endo/Exonuclease, localized in the cytoplasm [94,95]	-
				DIS3L2	3'-5' Endo/Exonuclease, localized in the cytoplasm [94,95]	Renal hamartomas nephroblastomatosis and fetal gigantism; Wilms tumor [105]
			Rrp6	hRRP6 (PM/Scl100, EXOSC10)	3'-5' Exonuclease [96]	Autoantigen in polymyositis, scleroderma, and PM/Scl overlab disease [98-100]
			Rrp47 (Lrp1)	C1D	Rrp6-interacting partner [97]	Autoantigen in PM/Scl overlab disease [106]

Table 2: Subunit homology in the *S. cerevisiae* and human nuclear exosomes. Yeast protein names are given in the first column. Human protein names are given in the second. Names used throughout this review are listed first, and alternative names used in the literature are listed in parentheses. A brief description of the role of the protein within the exosome is listed as well as any known links to disease in humans.

RNA substrates can be targeted by both endonucleolytic and exonucleolytic activities. It is proposed that the exonuclease activity is the major function of Dis3, and the endonuclease activity may function primarily to release stalled exosome substrates [80].

In vitro analysis and structural studies has suggested that Dis3 is also capable of degrading RNAs independently of the core Exo9 complex [64, 87, 90, 113]. Interestingly, isolated Dis3 has been shown to have increased exo-nucleolytic activity on certain substrates when it is not in a complex with the core barrel [90]. It was speculated that the structure of the core proteins partially inhibited Dis3 access to RNA substrates. In this context, it has recently been shown that there may be a second route for RNA digestion by Dis3 that is independent of the core [114]. This has lead to two distinct models for Dis3 degradation of RNA. The first model involves a substrate-specific degradation mechanisms in the RNA exosome with longer (>14 nt) single-stranded RNA substrates entering the exosome channel, inducing a conformational change in Dis3 so that the catalytic exonuclease site faces the base of the channel. When RNA substrates shorter than 12 nt were analyzed, the RNA is suggested to enter the Dis3 exonuclease site directly, without entering into the exosome channel and as a consequence did not induce a conformational change that was distinct from the isolated Dis3 structure.

Addition of Rrp6 in Yeast

Rrp6 is an RNase D family member that is also involved in mRNA decay *via* the exosome [96]. In *S. cerevisiae*, the Rrp6 exonuclease is only present in the nucleus, and it is hypothesized to hydrolyze RNA via metal ion catalysis [115]. Rrp6 contains an exoribonuclease domain (EXO), and HRDC domain and a C-terminal domain that associates with the Exo9 core (labeled Rrp6-CTD in Figure 4) [67, 116]. The EXO and HRDC domains have been shown to be the catalytic module through structural studies [117, 118]). However, it is not clear if the channel is used to direct RNA to the Rrp6 exonuclease.

Makino *et al.* solved a crystal structure of $Exo10^{Dis3}$ plus the C-terminal region of Rrp6 and an RNA duplex with a 31 nt 3' overhang [67]. To obtain this 2.8 Å resolution structure, the authors used the Exo9 core complex and a catalytically inactive Dis3 (D171N/D551N) mutant lacking both endo- and exonucleolytic activity (blue in Figure 4). Addition of the C-terminal region of Rrp6 was found to stabilize the interaction between the exosome and the RNA substrate. The structure of the Rrp6 C-terminal domain showed that it forms two regions that fold into the Exo9 core barrel: an α-helix that binds to Csl4 and an α-helix and β-hairpin that bind Mtr3 and Rrp43. The authors suggest that although this region does not bind RNA directly, it may play a role in stabilizing Csl4 and the path taken by the long RNA substrate through the barrel. Furthermore, it was noted that the addition of Rrp6 results in a rotation of the structure of Dis3 to closed conformation with the endonucleolytic PIN domain exposed to solvent. The duplexed RNA with 3' overhang allowed the capture of the exosome

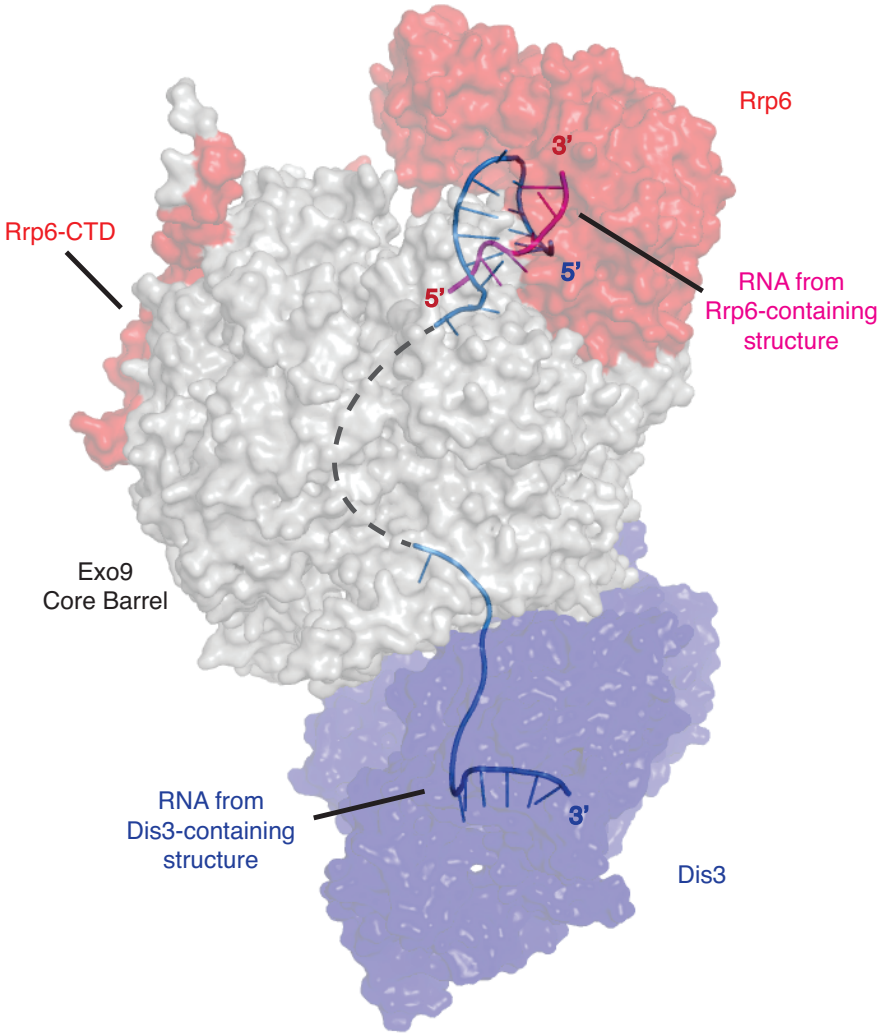

Figure 4: Composite structure of the nuclear exosome including Rrp6 and Dis3 exonucleases and RNA. This composite structure was modeled by overlaying figures containing the core barrel, a catalytically inactive Dis3 (Dis3(D171N/D551N)), an RNA duplex with a 31 nt overhang, and the CTD of Rrp6 [67] and the core barrel, Rrp6, and a 24 nt single-stranded poly(A) RNA [119]. The core barrel subunits are in gray. Dis3 and the path of the RNA from the structure containing Dis3 are in blue. Rrp6 and the path of the RNA from the structure containing Rrp6 are in red. A hypothetical path for the RNA through the core barrel is shown as a dashed line.

structure with single stranded RNA threaded through the barrel core. In addition, this crystallography arrangement allowed for melting and visualization of the most 5'- nt on the complementary strand at the entry pore formed by the cap proteins. These results suggest that the cap proteins serve to unwind RNA duplexes for entry into the core, analogous to proteins entering the proteosome. However, these ideas have not yet been confirmed *in vivo*. The RNA unwinding activities have been shown to be carried out by the Ski and TRAMP complexes that regulate exosome (or Exo10) function [120-123].

To better understand the role of Rrp6 in RNA degradation, Wasmuth *et al.* recently reported a crystal structure of a complex consisting of the Exo9 core, Rrp6, and 24 nt-long single-stranded poly(A) RNA molecule ($poly(A)_{24}$) [119]. The Makino *et al.* and Wasmuth *et al.* structures have been superimposed to make the composite structure shown in Figure 4. Wasmuth *et al.* previously proposed that Dis3 and Rrp6 use an overlapping channel within the core ring structure to interact with RNA substrates. Furthermore, it is also suggested that Rrp6 activity is modulated by its association with the Exo9 core complex, and Rrp6 stimulates Dis3-RNA binding and degradation [109].

Wasmuth *et al.* were able to obtain a 3.3 Å structure of the $Exo10^{Rrp6}$ complex (Exo9 core barrel plus Rrp6), using an Rrp6 (128-685) mutant lacking the PMC2NT domain (1-127), the last C-terminal residues, and exoribonuclease activity (D238N). In their structure, Rrp6 sits atop the S1/KH cap proteins of the Exo9 core (red in Figure 4). The $poly(A)_{24}$ RNA is positioned within the S1/KH ring with the active site of Rrp6 interacting with the 3'end of the RNA (pink in

Figure 4). This confirmation allows for the catalytic domain of Rrp6 to bind directly to the fourth-most 3'-nucleic acids in the poly$(A)_{24}$ chain, while the interactions between Rrp6 and the remainder of the RNA chain are non-specific to accommodate the degradation of any RNA sequence. The RNA substrate then passes through the Exo9 core channel, similar to that proposed for the Dis3 RNA substrates. In fact, the paths used by substrates for Rrp6 and Dis3 overlap within the S1/KH channel, with the RNA oriented in opposite directions to accommodate the location of the targeting exonuclease. Figure 4 illustrates the theoretical path marked in the core barrel by a dashed line. It was also proposed that the overlapping paths commit the exosome targets to degradation by either Rrp6 or Dis [119]. In the nucleus, interactions with cofactors such as Mpp6, Rrp47, and TRAMP are proposed to guide the RNA substrate toward a particular path, depending on the substrate and the required exosome function: editing, processing, or degradation.

Shared and Distinct Roles between Rrp6 and Dis3

Rrp6 is a distributive 3'-5' exonuclease (repetitive binding and release) and Dis3 is a processive 3'-5' nuclease [109]. Rrp6 can stimulate Dis3 binding and decay activities [109]. This may be due to a conformational change in the core that results in an increase in channel width when Rrp6 is bound. This theoretical mechanism of exosome regulation through channel gating in the exosome has also been seen in protein degradation in the proteasome (reviewed in [68, 124]).

Both Rrp6 and Dis3 appear to have some non-nucleolytic activity. Full deletions of either protein have more severe phenotypes than inactive mutants [87, 125]. This is supported by the structural data that indicates a requirement for specific fragments of Rrp6 and Dis3 for proper RNA exosome structural integrity and stability of their binding partners. A more recent study confirmed the independent functions of Rrp6 and Dis3 in a comprehensive manner. Gudipati *et al.* used high-resolution tiling arrays to study the effects of Rrp6 and Dis3 mutants on RNA degradation genome-wide [79]. The *RRP6* gene was fully knocked out, but catalytic mutations in both the endo- and exonucleolytic domains of Dis3 are lethal in yeast. Therefore, mutants were expressed in a doxycycline-repressible (Tet-*DIS3*) background strain. Mutants studied contained an inactivating mutation in the exonucleolytic domain (D551N), endonucleolytic domain (D171N), or both. Rrp6 and Dis3 mutants have similar effects at CUTs and SUTs, but distinct roles at other transcripts. Rrp6 appears to be the predominant nuclease in the processing of snRNAs while Dis3 plays a larger role in degrading tRNAs and intron-containing (i.e. unspliced) pre-mRNAs.

3.2. Other Cofactors in the Nucleus

As described above, Rrp6 and the nuclear exosome carry out a number of functions. The mechanisms underlying how a specific substrate is handle by the exosome remain unclear, but it is likely that proteins associated with the exosome aid in target/function specificity. Here, we discuss proteins associated with the nuclear exosome that aid in substrate recognition and Rrp6 activity.

Rrp47, aka Lrp1

Rrp6 interacts with the cofactor Rrp47 [97] (Figure 5). Rrp6 activity has been found to be at least partially dependent on Rrp47, but the impact of this interaction on the catalytic activity of Rrp6 are unclear. Rrp6 contains an N-terminal PMC2NT domain that interacts with Rrp47 [126]. Rrp47 can bind structured amino acids, and it is possible that Rrp47 facilitates Rrp6 activity by aiding in structured RNA recruitment [126]. It was suggested that this interaction is required to stabilize both Rrp6 and Rrp47 proteins [127-129]. Rrp6 expression is markedly decreased in yeast lacking Rrp47 when grown in minimal media, although this effect is not seen in rich media conditions [127]. Loss of Rrp47 has also been reported to decrease both transcription of *RRP6* and the stability of the Rrp6 protein. Interestingly, over expression of Rrp6 was found to suppress many of the phenotypes reported for Rrp47 mutants [127]. These results suggest that a primary function of Rrp47 is to stabilize and maintain sufficient Rrp6 expression levels. However, it has also been reported that Rrp47 may play roles in RNA processing and surveillance independent of Rrp6 [128]. Garland *et al.* developed a technique they termed DECOID (decreased expression of complexes by overexpression of interacting domains) in which they overexpress the Rrp6 N-terminal region shown to be the Rrp47-interacting domain to titrate Rrp47 out of Rrp6/Rrp47 complexes [128]. Once removed from Rrp6, the isolated Rrp47 was able to rescue synthetic lethal double mutants.

The N-terminal regions of Rrp6 and Rrp47 form a closely intertwined unit [130]. This structural unit was found to bind to the N-terminal region of the RNA helicase Mtr4, which can associate with the TRAMP complex to target RNA substrates to the exosome for degradation. Mutations in the N-terminal region of Rrp6, a portion that is important for Rrp47 binding, result in defects in 5.8S RNA processing *in vivo* suggesting that this intertwined structural unit is important for coordinating interactions with Mtr4 and proper substrate processing.

Mpp6

M-phase Phosphoprotein 6, or Mpp6, has also been shown to associate with the nuclear exosome [131] (Figure 5). Mpp6 was initially identified in a cDNA cloning screen performed in HeLa cells where it localized in the nucleus of interphase cells [132]. Schilders *et al.* found that human Mpp6 was important for the proper processing of the 3' ends of 5.8S rRNAs [133]. In 2008, Milligan *et al.* identified the homolog in *Saccharomyces cerevisiae* [134]. The yeast homolog of Mpp6 was found to be an RNA-binding protein required for proper surveillance of pre-mRNAs, pre-rRNAs, and the degradation of cryptic non-coding RNAs transcribed from ribosomal DNA spacer heterochromatin [134]. It has been suggested that Mpp6, and probably other cofactors like it, play a role in specifying the function to be carried out by the exosome regarding specific substrates [59].

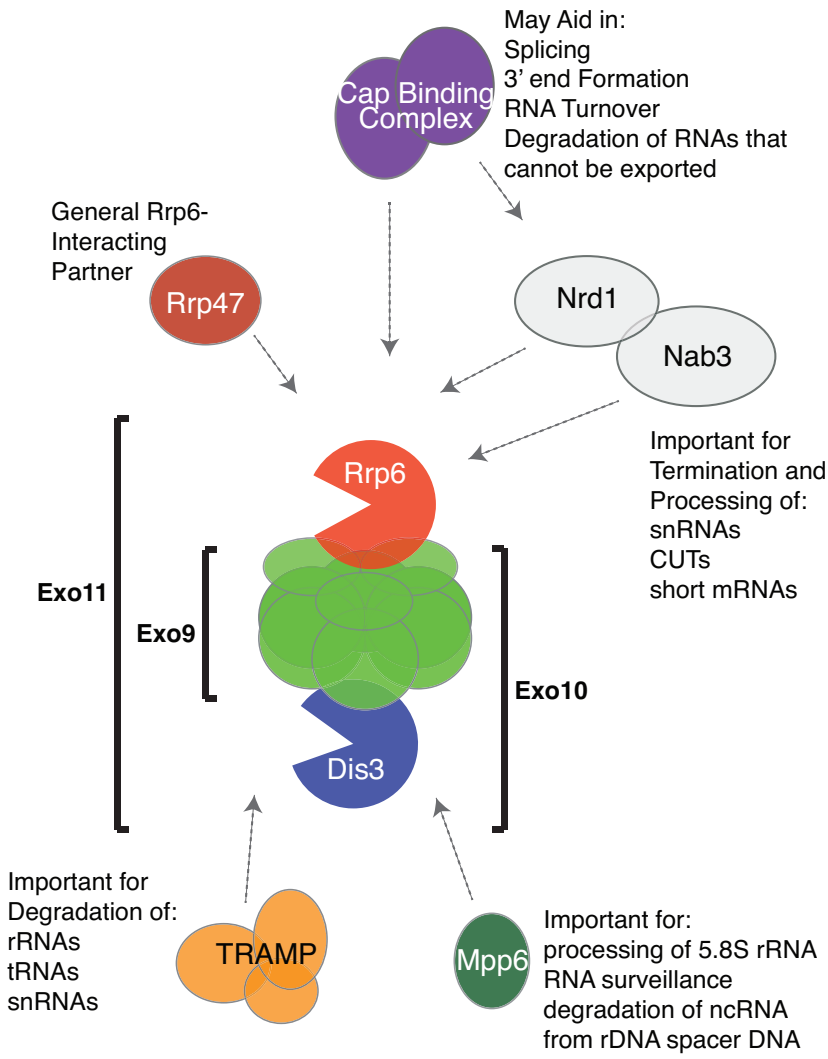

Figure 5: Schematic representation of the *S. cerevisiae* nuclear exosome and its known cofactors. The core barrel structure (also known as Exo9), in green, consists of nine subunits: Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, and Mtr4 (the six RNase PH-like proteins), and Rrp4, Rrp40, and Csl4 (the three S1/KH cap proteins). The nuclease Dis3 joins the core barrel in both the cytoplasm and the nucleus. This complex is also called the (Exo10). Rrp6 only joins the core exosome in the cytoplasm, in a complex also known as Exo11. The cofactor Rrp47 (dark red) binds to Rrp6 and structured RNA and improves the stability of Rrp6. Nrd1 and its binding partner Nab3 (gray) interact with Rrp6, both together and independently of one another to coordinate termination and processing of short (<1000 nt) RNAs. The Cap Binding Complex (purple) interacts with Rrp6 during co-transcriptional processes. The Cap Binding Complex also interacts with Nrd1 and Nab3, which may lead to an alternative, indirect interaction with Rrp6. Mpp6 (dark green) is a general nuclear exosome cofactor involved in a number of exosome-dependent mechanisms such as processing and degradation of a RNA arising from rDNA arrays.

The TRAMP Complex

In the nucleus, the RNA exosome interacts with an RNA helicase, Mtr4, and an associated complex known as TRAMP (the Trf4/5, Air1/2, Mtr4 Polyadenylation Complex, reviewed in [66]) (Figure 5). Mtr4 can function alone to unwind RNA substrates or it can act as part of the TRAMP complex [135, 136]. Mtr4 contains a 'ratchet helix' and arch domain required for RNA processing, although the specific roles for each of these domains in Mtr4 function remain unclear [137-139]. The TRAMP complex aids in substrate specificity of the exosome [140], and prepares some target RNAs for processing by the nuclear exosome by adding a short (3 – 50 nt) poly(A) tail which is thought to provide an unstructured starting point for the exosome to thread the substrate RNA through the core barrel of the exosome [60, 71, 122, 141]. This poly(A) tail is added by one of two poly(A) polymerases in the TRAMP complex, Trf4 or Trf5. Trf4 and Trf5 exist in separate complexes, called TRAMP4 and TRAMP5, respectively. The TRAMP complex has been shown to be important for degradation of pre-tRNAiMet [142, 143] as well as for the decay of rRNA and small nuclear/nucleolar RNAs [71, 122, 141].

3.4. Roles of the Nuclear Exosome in Transcription

Rrp6 May Have Some Function Apart From the Core Exosome

The TRAMP complex and core exosome barrel help to target specific RNA degradation substrates to Rrp6 and Dis3, but there is evidence that Rrp6 may be capable of performing some of its functions apart from the core exosome.

Callahan *et al.* studied the effects of disrupting, individually or in combination, the exosome subunits Rrp6, Dis3, Rrp43 [116]. While the individual mutants each significantly altered the transcriptome, there were changes in distinct classes of transcripts. Disruption of Rrp6 and Dis3 in combination resulted in a synergistic increase of the Dis3- and Rrp43-specific transcripts. The expression of the Rrp6-specific transcripts remains the same when *RRP6* is deleted, whether or not Dis3 or Rrp43 are present. These findings suggest that Rrp6 may have exosome-independent functions that are therefore, not dependent on Dis3 and Rrp43. In support of this hypothesis, they also found that an Rrp6 mutant that disrupts the interaction between Rrp6 and the core exosome can carry out 3' end processing of 5.8S rRNA and snRNAs. However, this mutant is no longer capable of degrading transcripts shown to require both Dis3 and Rrp6 function. This suggests that specific Rrp6-dependent activities require interaction with the exosome, and some activities do not. This may prove to be important in understanding the role of Rrp6 in RNAPII transcription termination.

Interactions with Nrd1

Rrp6 has been found to interact with the RNAPII transcription termination factor Nrd1 by immunoprecipitation [70] (Figure 5). Nrd1 interacts with Nab3 and Sen1 as well as the Cap Binding Complex (CBC) and is responsible for RNAPII termination of short RNA target genes (<1000bp) [4, 44, 144-146] including snRNAs and CUTs [41, 147]. As noted above, Nrd1 binds Ser5-P in the CTD of RNAPII through the Nrd1 CID, as well as consensus sequences in the nascent

RNA through an RNA Recognition Motif (RRM Domain) in the Nrd1 protein. Because Nrd1 is a sequence-specific RNA binding protein, it has been suggested that Nrd1 binding may block Rrp6-dependent degradation from proceeding past specific sequences [70]. This could be a mechanism to allow Rrp6 and the nuclear exosome to distinguish between RNA targets for degradation or those for 3' end processing. Intriguingly, it has been shown that misprocessed transcripts in yeast caused by ectopic expression of the bacterial RNA Polymerase termination factor Rho are degraded by Rrp6 and the nuclear exosome [148], and the formation of these transcripts results in increased recruitment of Nrd1 to the transcription machinery by the Nrd1 CID and RRM domain [56]. Association of Rrp6 was also increased at these complexes. This, taken in context with the known interaction between Rrp6 and Nrd1, suggests that Nrd1-interaction with RNAPII and the nascent RNA may be an important step in recruitment of Rrp6 and the nuclear exosome to the RNA substrate. In support of this model, Heo *et al.* has shown that disrupting the normal Nrd1-RNAPII CTD interaction by replacing the Nrd1 CID with the CID derived from Rtt103 that recognizes Ser2-P RNAPII CTD, reduces binding of the chimeric Nrd1 to Rrp6 [45].

The Nrd1 binding partner, Nab3, may also interact with Rrp6 independently of Nrd1 [149]. Fasken *et al.* recently found that overexpression of Nab3 in strains carrying mutations in the TRAMP RNA-binding subunits Air1 and Air2 can suppress the slow growth phenotype and decrease the amount of extended transcripts. This effect was not observed for Nrd1 or Sen1 and does

not require the Nrd1-interacting domain of Nab3, but it does require the RNA-binding domain of Nab3 and Rrp6. Deletion of the Nrd1 binding domain of Nab3 revealed that Nab3 is capable of interacting with Rrp6 independently of Nrd1 [149].

3.5. Transcription Termination

Gene Regulation by Early Termination

Studies in human cells suggest that Rrp6 and SetX (the human homolog of Sen1) may cooperate with Microprocessor to regulate RNAPII elongation via RNAi at some genes [150]. The Microprocessor complex is required for processing of functional microRNAs and includes the RNase II Drosha and the dsRNA-binding protein Dgcr8 [151]. Typically, RNA regulates protein expression post-transcriptionally by binding to the target mRNA and inducing RNA degradation or inhibition of mRNA translation. However, Wagschal *et al.* reported that cooperation between Microprocessor, Rrp6, and Setx regulate *HIV-1* transcription by inducing premature termination [150]. The authors propose a model in which Microprocessor is recruited to the nascent stem-loop TAR RNA located in the most 5' portion of the *HIV-1* gene, triggering RNAPII pausing and subsequent cleavage of the RNA by Rrp6. Studies in yeast suggest that this mechanism of promoter-proximal pausing followed by premature termination induced by cooperation between Microprocessor and Rrp6 and Setx and/or other NNS-dependent termination factors may function to regulate transcription elongation of a number of genes [44, 47, 152, 153].

3.6. Nuclear RNA Processing and Surveillance

With recent advances in deep sequencing technology, it has become obvious that the majority of the genome can be transcribed into RNA. Rrp6 and the nuclear exosome have been found to have multiple roles in regulating the fate of these transcripts. The exosome must identify and properly respond to signals that delineate which transcripts are to be completely degraded, which should have their 3' ends partially degraded and precisely how far, and which should avoid the activities of the exosome altogether. A summary of some of known RNA processing and degradation activities of Rrp6 and the nuclear exosome is shown in Figure 6.

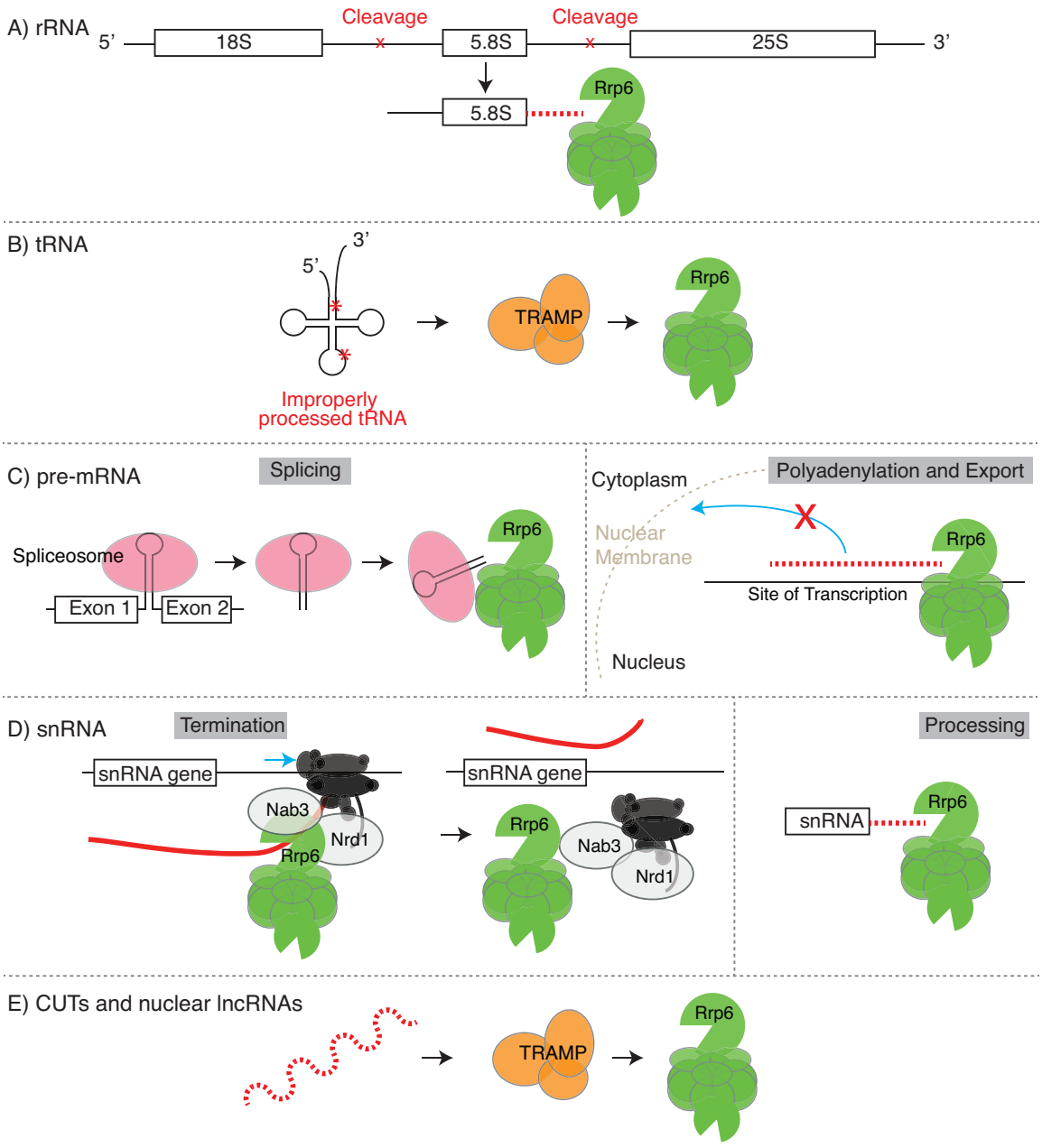

Figure 6: Multiple mechanisms by which Rrp6 processes or degrades its many RNA substrates. (A) The 18S, 5.8S, and 25S rRNAs are transcribed as one molecule and cleaved and processed by several nucleases in the sequence of events leading to the rRNA that are incorporated into ribosomes. Specifically, Rrp6 is required for proper processing of the 3' end of the pre-5.8S product. (B) Improperly processed tRNAs in the nucleus are targeted by the TRAMP complex which adds a short polyA tail to the pre-tRNAs for targeting to the exosome for complete degradation by Rrp6. (C, left) Exo11 interact directly with the

spliceosome to degrade introns co-transcriptionally. (C, right) Rrp6 also degrades mRNAs that cannot be exported from the nucleus and accumulate at the transcription site as a consequence of improper processing 3' ends or inhibition of nuclear export of the mRNA. (D, left) Rrp6 is required for proper transcription termination by Nrd1. The mechanism by which Nrd1 causes RNAPII termination is not known. It may require an Nrd1 interaction with Rrp6 to properly terminate transcription by a process involving release of DNA and RNA from RNAPII and the termination factors. (D, right) Rrp6 processing of the 3' end of snRNAs terminated in a heterogeneous "zone." The extended 3' ends of the pre-snRNA are trimmed back by Rrp6. (E) Cryptic Unstable Transcripts (CUTs) and unstable nuclear lncRNAs are both targeted to the Exo11 by the TRAMP complex.

rRNA

Rrp6 and the other 10 subunits of the RNA exosome are required for proper 3' end processing of the 5.8S rRNA [96, 133, 135, 154, 155] (Figure 6A). In eukaryotes, rRNA precursors (all but the 5S precursor) are transcribed by RNA Polymerase I from an array of precursor units that contain the 18S, 5.8S, and 25S products separated by transcribed spacer regions. Each unit is transcribed as one molecule, and hundreds of these units may exist in the genome. These large precursor transcripts are then processed by a series of cleavage and nucleolytic reactions. The 3' end of the 5.8S precursor is heterogeneous after cleavage separating it from the 25S precursor. The exosome processes the 3' end to a specific length resulting in a functional 5.8S rRNA [154]. Furthermore, depletion of individual subunits of the exosome each results in inhibition of early pre-rRNA cleavage [156]. None of the affected cleavage steps require 3' – 5' degradation, but the exosome degrades improperly processed rRNA precursors that form following the inhibition of rRNA processing factors [156].

The 5S rRNA is transcribed by RNAPIII, and has been shown to be targeted by the TRAMP complex when the 5S rRNA is improperly processed [157]. Using the CRAC technique to identify RNA-protein binding sites, Wlotzka *et al.* identified RNA transcripts bound to Rrp6, the Trf4 component of the TRAMP complex, and Nrd1 and Nab3 that were required for termination of short RNAPII transcripts [60]. This study found that Nrd1, Nab3 and Trf4 crosslinked to 5S rRNA sequences that often contained a short poly(A) tail, suggesting that these transcripts were intermediate forms of exosome-degradation targets.

tRNA

Rrp6 and other components of the exosome and TRAMP complexes have also been shown to play important roles in 3' end processing and turnover of tRNAs transcribed by RNAPIII [141-143, 157-159] (Figure 6B). Pre-tRNAs contain 5' and 3' leader sequences, and a few tRNA genes have introns. During processing, these sequences are removed, a 3' CCA tail is frequently added, and specific base modifications are made. The TRAMP complex is suggested to function in the recognition of improperly processed tRNAs (reviewed in [36]). The study initially identifying the TRAMP complex found that Trf4 preferentially targeted misfolded tRNAs while avoiding properly processed tRNAs [141]. As with other targets, TRAMP adds a short poly(A) tail to the RNA, targeting it for degradation by Rrp6 and the exosome.

In the CRAC study by Wlotzka *et al.*, Nrd1, Nab3, and Trf4 were found to bind to many pre-tRNA sequences containing 5' and 3' leaders and introns [60]. However, almost none of the bound pre-tRNAs contained the 3' CCA tail but did have a short poly(A) tail. The authors suggest that pre-tRNA sequences are targeted by the TRAMP/exosome system after improper processing. An RNA-binding protein called La binds to the 3' end of all RNAPIII transcripts protecting them from degradation. It has been found that La-binding can protect mildly defective pre-tRNAs from degradation by Rrp6 [158], and it is perhaps the interplay between La and the TRAMP complex that allows for the specific targeting of improperly processed tRNAs.

sno/snRNAs

Rrp6 activity and the Rrp47 cofactor are required for 3' end processing of sn/snoRNAs. These snRNAs are highly transcribed and abundant non-coding RNAs that facilitate rRNA processing. Some snRNAs are transcribed as part of a larger transcript and subjected to processing, originating either from an intron of a protein-coding gene, or as a polycistronic transcript containing several snRNAs. The snRNAs originating from larger transcripts must be processed at the 5' and 3' ends by a number of different exonucleases [160-164]. For example, Rrp6 and its cofactor Rrp47 are required for proper 3' end trimming of some of these snRNAs [165, 166].

Other snRNAs are transcribed individually, in genes with their own promoter and terminator. These individually transcribed snRNAs are terminated by the NNS pathway [70] (Figure 6D). As mentioned above, Nrd1 is thought to terminate within a non-discrete zone dependent on the kinetic interaction between Sen1 and RNAPII [51]. Termination within a zone results in heterogeneous 3' transcript ends that are processed by Rrp6 [165, 166]. As with other exosome targets, the polyadenylation of pre-snRNAs by the TRAMP complex aids in targeting snRNAs to the exosome [58]. Interestingly, the poly(A) binding protein Pab2 may also play a role in processing snRNAs in yeast [167]. It has been suggested that Nrd1 binding at specific sites on the 3' end of pre-snRNA transcripts may signal termination of processing by the exosome, resulting in homogeneous 3'ends [45, 70].

CUTs

Increased sensitivity of microarrays and sequencing technologies have uncovered that much of the eukaryotic genome can be transcribed into RNA. Many non-coding RNAs can be stabilized in the absence of Rrp6, suggesting that these RNAs are frequently synthesized and rapidly degraded in wild-type cells. Multiple reports have used tiling microarray technology to identify novel exosome targets by knocking out *RRP6*. [71, 72, 147, 153, 168, 169] (Figure 6E). Many CUTs originate from intergenic regions previously thought to be untranscribed [71], especially around promoter regions of protein coding genes [75]. Like other targets of Rrp6, these cryptic transcripts are targeted for degradation by subunits of the TRAMP complex. In 2009, it was found that many of these cryptic transcripts were initiated at bidirectional promoters in nucleosome free regions [72, 73]. Many of these bidirectional promoters correlate with promoters of protein-coding genes where the CUTs were transcribed in the divergent orientation from the opposite strand. Other CUTs found to begin upstream of the mRNA start site, antisense to the protein-coding gene and overlapping the 5' end, have been implicated in regulation of mRNA expression [41, 152, 170, 171]. These studies and the discovery that the Nrd1 termination pathway is involved in CUT termination [147] have led to a transcription attenuation model in which transcription and rapid degradation of non-coding RNAs result in the differential expression of neighboring protein-coding genes. Interestingly, it was recently shown that Rrp6 plays a greater role in degrading these antisense CUTs present

within protein-coding genes than CUTs originating from bidirectional promoters [172].

lncRNA

Long noncoding RNAs (lncRNAs) are defined as RNAs longer than 200 nt with no detectable protein-coding potential [173]. Many transcripts that fall into this category have no known function as of yet, but others have been shown to play important roles in gene regulation, both transcriptionally and post-transcriptionally [174, 175]. Many of these functional lncRNAs are capped and polyadenylated analogous to protein-coding mRNAs [176-178]. Expression of some lncRNAs is thought to be regulated by Rrp6-dependent degradation (Figure 6E). Deletion of *RRP6* results in an increase in the half-life of specific lncRNAs [179]. However, most lncRNA transcripts escape nuclear degradation and are instead exported to the cytoplasm. There, they can be decapped and degraded to regulate their activity [180]. It has been suggested that lncRNAs are transcribed from 10- to 20-fold more genomic sequence than protein-coding RNAs [175], and Rrp6 surveillance may be an important quality control step in lncRNA function.

4. Regulation of CTD Phosphorylation by Rtr1

Our lab and others have shown that Rtr1 is a novel regulator of CTD phosphorylation that dephosphorylates Ser5 during early transcription elongation in *S. cerevisiae* [8, 181-184]. Deletion of *RTR1* results in extension of Ser5-P

levels as RNAPII proceeds toward the transcription termination site. In some cases extended Ser5-P correlates with improper transcription termination, yielding RNA transcripts with 3' ends that extend beyond the proper Transcription Termination Site (TTS). Such extended transcripts can include extended 3' untranslated regions (UTR) as well as parts of the 5' UTR and coding regions from downstream genes [8]. These findings support the hypothesis that after deletion of *RTR1*, the altered pattern of phosphorylation across the RNAPII CTD disrupts the function of the termination machinery. However the precise mechanisms by which Rtr1 regulates the RNAPII termination machinery are yet to be determined.

4.1. The Yeast Serine 5 Phosphatases: Rtr1 and Ssu72

As highlighted above, two known phosphatases in *S. cerevisiae*, Rtr1 and Ssu72, remove Ser5-P at different stages of the RNAPII RNA transcription cycle (Figure 2). Rtr1 is an atypical phosphatase required for proper removal of Ser5-P from RNAPII during early elongation and removes the anti-termination mark, tyrosine 1 phosphorylation [8, 17]. Rtr1 and its human homolog, RPAP2, are both capable of removing the Ser5-P mark, but not Ser2-P or Ser7-P [8, 181, 183]. *RTR1* is nonessential in budding yeast. Deletion of Rtr1 in yeast, or knockdown of RPAP2 (via siRNA) in human cells results in the spread of Ser5-P RNAPII into the 3'-end of RNAPII target genes, including *PMA1*, *ACT1*, and *PYK1* in yeast and β-actin and the U2 snRNA gene in humans [8, 183]. Rtr1 interacts primarily with RNAPII, and deletion of nonessential RNAPII subunits Rpb4 or Rpb9 display

negative synthetic genetic interactions with deletions of the *RTR1* gene [9, 185]. In addition, *RPB9*, *RPB7*, and *RPB5* can act as high-copy suppressors of *RTR1* knockout phenotypes [185].

Affinity purification mass spectrometry data has revealed that recruitment of Rtr1 to RNAPII requires the Ser2 kinase Ctk1 *in vivo*, and Rtr1 recognizes a specific phosphorylated form of RNAPII not recognized by the other Ser5 phosphatases Fcp1 and Ssu72 [9]. Of interest, RPAP2 has also been shown to require other CTD-binding proteins for its recruitment to dephosphorylate Ser5. The proteins RPRD1A, RPRD1B, and RPRD2 bind RPAP2 and Ser2- and Ser7-phosphorylation marks [184, 186]. RPRD1A and RPRD1B act as a scaffold to position RPAP2 on the CTD and have been shown to facilitate RPAP2 activity [184]. There are no RPRD homologs in *S. cerevisiae* or *Kluyveromyces lactis.* The precise catalytic mechanism of Ser5 dephosphorylation by Rtr1 is yet to be determined. Structural studies on Rtr1-substrate complexes to date have been unsuccessful, and the *K. lactis* Rtr1 that have been crystalized were found to be inactive or showed low activity *in vitro* [17, 187].

The phosphatase Ssu72 removes both Ser5 and Ser7 phosphorylation in budding yeast and appears to be involved in multiple stages of RNAPII transcription [182, 188-191]. Ssu72 is essential in budding yeast but is dispensable for viability in the fission yeast *Schizosaccharomyces pombe* [192]. Ssu72 has been shown to be a subunit of the CPF complex responsible for proper termination of polyadenylated RNAs and snRNAs [15, 193]. Within the CPF complex, Ssu72 interacts closely with Pta1 in yeast or its human homolog

Symplekin [12, 194]. Ssu72 activity is promoted by the CTD prolyl isomerase Ess1 that changes the conformation of the CTD to one more favorable to Ssu72 [63]. Ssu72 activity is also affected by adjacent Thr4-P, the presence of which decreases Ser5-P dephosphorylation on a CTD peptide 4-fold [195]. It remains unclear why there are two Ser5 phosphatases with seemingly overlapping functions, but it is likely that the specific activities of the phosphatases are modulated by the complex variations of post-translational CTD modifications. Additionally, the RNAPII subunits Rpb4/7 may help recruit Ssu72 [196], while Rtr1 interacts with a 10-subunit form of RNAPII lacking Rpb4/7 as characterized by mass spectrometry [9, 197].

4.2. Effect of the Loss of Rtr1

Deletion of *RTR1* results in continued Ser5 phosphorylated RNAPII toward the 3'-end of target genes and in many cases into or past the 3'-UTR [Hunter, G.O. and Mosley A.L., unpublished data]. An example of these data is illustrated in Figure 7. To study genome-wide the extent to which Ser5-P persists in to the 3' end of transcripts, we performed Chromatin Immunoprecipitation (ChIP) using antibodies against Ser5-P RNAPII CTD peptides and identified the regions of the genome immunoprecipitated by high-density microarray analysis. The microarray allows for the identification of regions with Ser5-P RNAPII in the 5'- and 3'-ends of genes independent of the genes length using multiple probes per gene (resolution is 50 nt). In wild-type strains, Ser5-P RNAPII peaks shortly after transcription elongation begins and gradually decreases along the length of the

gene until it rapidly decreases at the TTS. However, when *RTR1* is deleted, Ser5-P remains high throughout the length of the gene and past the TTS into the intergenic region. The change in Ser5-P distribution in the absence of *RTR1* is shown in black (Figure 7).

We previously observed that deletion of *RTR1* results in improper RNAPII termination at *NRD1* [8]. This defect in transcription results in a product with 3'-ends that extend beyond the proper TTS and includes the 3'-UTR and part of the adjacent gene *MRPL17* [8]. Read-through transcription was also observed in human cells at the U2 snRNA gene followed knock down of RPAP2 [183]. Taken together with the global persistence of Ser5-P into the 3'-ends of transcribed regions in downstream genes, these findings support the hypothesis that after deletion of *RTR1* alters the pattern of phosphorylation across the RNAPII CTD, disrupting RNAPII function and/or recruitment of the termination machinery. The mechanisms through which the termination machinery is disrupted in the absence of Rtr1 are yet to be determined.

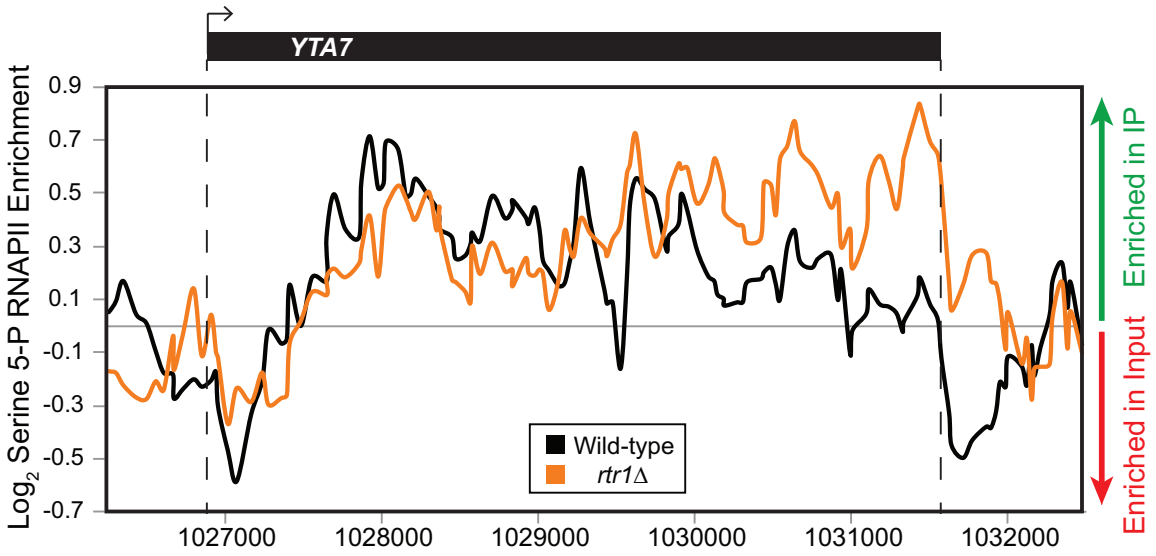

Figure 7: Ser5-P RNAPII increases across transcribed regions in strains with *RTR1* deleted. Ser5-P pattern measured by Chromatin-immunoprecipitation followed by microarray. The Ser5-P CTD profile shown is for *YTA7*, a moderately transcribed RNAPII target gene. The figure is drawn to scale with the transcription region for *YTA7* at the top. The transcription start site (TSS) and transcription termination site (TTS) are also labeled as determined by RNA-Seq by [35]. The distribution of Ser5-P RNAPII across *YTA7* in wild-type cells is shown in blue. Ser5-P RNAPII is most abundant at the 5'-end of the transcribed gene and begins to drop off steadily across the gene as transcription proceeds. Levels of Ser5-P are the lowest just proximal to the TTS and then are sharply lower distal to the TTS. The altered distribution of Ser5-P RNAPII across *YTA7* in *rtr1Δ* cells is shown in black. The abundance of Ser5-P in the 5'end is similar to that in wild-type cells, but it does not decrease as transcription of the gene proceeds. As illustrated in the diagram, Ser5-P continues into the 3'end of the gene, and even extends into the intergenic space that is 3'- to the transcribed gene.

METHODS

1. Yeast Strains

All yeast strains used are isogenic to BY4741. *RRP6* deletion strain is from the yeast knockout collection (Open Biosystems) [198]. *RTR1* was knocked out of wild type (WT) and the *RRP6* deletion strain by homologous recombination with a kanamycin cassette to create the *RTR1* deletion and *RTR1/RRP6* double deletion strains. The Rpb3-FLAG WT strain was produced by yeast transformation and homologous recombination with a tagging cassette containing sequences for a 3X-FLAG tag upstream of the *URA3* gene from the plasmid pBS1539 (all primer sequences available upon request) [199]. The Nrd1-TAP strain is from the yeast TAP-tag collection (Open Biosystems). The Rpb3-FLAG and Nrd1-TAP *RTR1* deletion strains were made by amplification of the *RTR1* knockout cassette from the *RTR1* deletion strain and transformation into the wild-type Rpb3-FLAG and Nrd1-TAP strains respectively. All transformations were confirmed by PCR genotyping. FLAG-tagged strains were made by homologous recombination using plasmids obtained from Funakoshi and Hochstrasser [200] to insert the 3xFLAG tag DNA sequence from the pFA6a plasmid into the genome at the 3'-end of the *SSU72* gene, just prior to the stop codon, in both WT (BY4741) and *rtr1Δ* strains (Open Biosystems). All TAP-tagged proteins were previously described and obtained from the Open Biosystems TAP-tagged library (http://www.openbiosystems.com).

2. Protein Purification and Proteomics Analysis

2.1. Protein Complex Identification by Affinity Purification

Cells were grown overnight in YPD to OD600 ≅ 1.5 and then collected by centrifugation for 10 minutes at 4000 x g,. Cell pellets were washed in H_2O and resuspended in in 25mL TAP lysis buffer per 2.5 grams of pellet (40mM Hepes-KOH, pH 7.5; 10% glycerol; 350mM NaCl; 0.1% Tween-20; fresh yeast protease inhibitors (diluted to 1X) and 0.5mM DTT). The cells were slowly transferred to liquid nitrogen using a syringe. The frozen cells were pulverized with a mortar and pestle and lysed further in a Waring Blender with dry ice. The frozen lysate was transferred to a new container and allowed to thaw at room temperature. The resulting extract was treated with 100units DNase I and 10 μ L of 30mg/mL heparin for 10 minutes at room temperature and clarified by centrifugation as previously described [201, 202].

To purify Ssu72 and interacting proteins, the lysate was incubated with FLAG-agarose resin at 4^{o}C overnight. The resin and bound proteins were removed from the lysate by gravity flow through a 30mL Bio-Rad Econoprep column and washed on the column with 60mL TAP lysis buffer. The resin was resuspended 300µL of 50mM ammonium bicarbonate pH 8.0 and transfer to a microcentrifuge tube for on bead digestion with 5µL of Trypsin Gold (0.1µg/µL) overnight with shaking at 37°C. The supernatant containing the digested proteins was removed and treated with 20µL of 90% formic acid to inactivate the trypsin. The samples were separated into 3 aliquots for technical replicate LC/LC-MS/MS analysis.

2.2. MudPIT-LC/MS Analysis

MudPIT (Multidimensional Protein Identification Technology) is an LC/LC-MS/MS technique that uses two-dimensional chromatography to separate complex peptide mixtures by both charge and hydrophobicity. Each sample was loaded onto a two-phase column containing strong cation exchange resin, which binds positively charged ions, and reverse phase C18 resin, which will retain peptides based on their hydrophobicity. The samples were eluted off the column by the MudPIT protocol of 10 steps of increasing salt concentrations (50-350mM ammonium acetate) followed by an organic gradient (20-80% acetonitrile). All chromatography solutions also contained 1% formic acid.

During each of the 10 MudPIT steps, peptides were eluted into a ThermoFisher LTQ Velos for mass spectrometry (MS) analysis. Raw spectrum data from the MS analysis were submitted for protein identification by Proteome Discoverer software (Thermo) version 1.3 using SEQUEST as the database search algorithm. Database searches were performed against a FASTA database from the National Center for Biotechnology Information (NCBI) containing *Saccharomyces cerevisiae* protein sequences and sequences for common contaminants. Reversed versions of all non-redundant proteins were also searched and used to calculate False Discovery Rates (FDR). Hierarchical clustering was performed as previously described [197, 201]. The most highly enriched proteins were clustered using the Normalized Spectral Abundance Factor (NSAF) values calculated as previously described.

3. Analysis of RNA

3.1. RNA Isolation

RNA was extracted using the hot acid phenol method. Strains were grown in 100ml YPD medium to an OD_{600} of 0.8. Cells were collected by centrifugation, washed, and resuspended in 10ml AE buffer (50mM sodium acetate at pH 5.2, 10mM EDTA) in a Nalgene™ phenol-resistant 50ml tube. Eight hundred microliters of 20% SDS and 10ml cold acid phenol were added to each sample and mixed well by vortexing. Samples were incubated at 65^{o}C for 10 minutes with vortexing every minute then cooled on ice for 5 minutes. Samples were centrifuged for 15 minutes at 10,000 rpm. The top phase was transferred to a pre-spun 50ml 5 PRIME ™ Phase Lock Gel tube (Ref # 2302870). A total of 13ml of chloroform was added and well mixed before centrifuging for 10 minutes at 3000 rpm. The top phase was poured into a new phenol-resistant tube, and 1/10 volume sodium acetate at pH 5.2 and equal volume room temperature isopropanol was added. The precipitated RNA was collected by centrifugation for 45 minutes at 12,000 rpm. The pellet was washed with 70% ethanol, allowed to dry in a fume hood, and resuspended with molecular biology grade water. The Ambion DNase-turbo kit was used to degrade any contaminating DNA. The quality of the samples was determined with an Agilent Bioanalyzer before preparation of the sequencing libraries.

3.2. SOLiD5500xl sequencing methods

Standard methods were used for RNA-Seq library construction, EZBead preparation and Next-Gen sequencing, based on Life Technologies SOLiD 5500xl system. Briefly, RNA quality was assessed with on an Agilent Bioanalyzer. Each sample was loaded onto a lane on a Bioanalyzer chip and separated by size with electrophoresis. The separated fractions result in peaks detected by the analyzer, and these peaks were compared to the expected size and intensity of RNA peaks from *S. cerevisiae*. The Bioanalyzer software compared the experimental peaks to the reference peaks to calculate an RNA Integrity Number (RIN) to quantitatively represent the quality of the RNA on a scale of 1-10, in which a higher number represent higher quality RNA.

Five microgram of total RNA per sample (RIN equal or higher than 6.0 by Agilent Bioanalyzer) was applied in library preparation. Large ribosomal RNAs (rRNAs) were first depleted using the standard protocol of RiboMinus™ Transcriptome Isolation Kit for yeast (Ambion, Cat# K1550-03), and rRNA-depleted RNA was concentrated with the PureLink RNA Micro Kit (Invitrogen, Cat# 12183-016) using 1 volume of Lysis Buffer and 2.5 volumes of 100% ethanol. Following rRNA depletion, whole transcriptome library was prepared and barcoded per sample using the standard protocol of SOLiD Total RNA-Seq Kit (Life Technologies, Cat# 4445374). Each barcoded library was quantified by quantitative PCR using SOLiD Library Taqman qPCR Module (Life Technologies, Cat#A12127), and pooled in equal molarity. Fifty microliters of 500pM of pooled library was used in subsequent EZBead preparation, which involves bead

emulsion, bead library amplification, and bead enrichment using Life Technologies EZ Bead™ E120 System (Cat# 4465571). Useful beads were amplified from one fragment, resulting in the same sequence repeated across the entire bead. Approximately six hundred million enriched beads then were deposited onto each lane of a 6-lane SOLiD 5500xl flow chip.

Finally sequencing by ligation was carried out using standard single-read, 5'-3' strand-specific sequencing procedure (75 bp-read) on SOLiD 5500xl Sequencer. Specifically, 8 nt probes with one of 1,024 possible sequences bound on the 5'-end to one of four possible dyes corresponding to the two 3'-most nucleotides are used to determine the sequence of each fragment. First, primers were hybridized to the P1 adaptor on every fragment. A probe complementary to the fragment sequence is ligated to the adaptor primer. The fluorescent dye corresponding to the first two nucleotides is read by the sequencer, then the three 5'-most nucleotides and the adaptor are cleaved and released. The sequencing reaction is reset and repeated. In this way, two nucleotides are detected, separated by three nucleotides of undetected sequence. By overlapping multiple fragments per bead (amplified from the same fragment), each nucleotide is sequenced twice per fragment.

3.3. RNA Sequencing Alignment

The resulting 75 bp solid reads were mapped to the *Saccharomyces cerevisiae* reference genome sacCer3 using in-house mapping pipelines that utilizes bfast-0.7.0a [203]. Briefly, using our RNA-Seq pipeline, poor quality and rRNA/tRNAs reads were first discarded. The remaining reads were mapped to

reference genome sacCer3 and a splice-junction library, respectively; the genomic and splice-junction library mapping were merged at the end. In a different pipeline, the rRNA/tRNAs were kept and the reads were mapped to the reference genome sacCer3 since in yeast there is some splicing but most protein coding genes do not have introns. Read counts were calculated using bamutils from NGSUtils [204].

Differential gene expression was analyzed using edgeR [205]. Gene expression was quantified using the Reads Per Kilo-base per Million mappable reads (RPKM) equation $R= (10^9C/NL)$, where R is the expression of the gene, C is the number of reads that fall within the gene annotation, N is the depth of sequencing, and L is the length of the gene. The edgeR statistical package assumes a negative binomial distribution of the data. All raw and processed files from the RNA sequencing performed for the *RRP6* deletion study have been deposited to Gene Expression Omnibus [GEO] under the accession number GSE57155. Annotations for transcripts that showed significant changes in *rrp6Δ* cells were used for subsequent differential expression analysis to generate the final dataset**.**

3.4. Manual annotation of novel transcripts

Following data alignment, snRNA transcripts were manually inspected individually using the Integrative Genomics Viewer [206, 207]. For all snRNAs in a tail-to-tail orientation with a downstream gene, the snRNA-ET annotation started just after the end of the snRNA annotation until continuous reads on the same strand were no longer detected. For all snRNAs in a tail-to-head

orientation, the snRNA-ET annotation started just after the end of the snRNA annotation and was ended just prior to the 5' end of the annotation for the downstream gene. Annotations for ET for snRNAs that were encoded within introns were ended just prior to the 5' end of the exon for the parent transcript. To identify antisense transcripts with significant changes in differential expression, the strand was reversed for all sense annotations for the coding region of each ORF-Ts and the text "AS_" was added in front of the ORF-T name. The annotations for the 5' and 3' UTR were not included. We also included an annotation for the transcript upstream of *IMD2*, known to be terminated by the NNS pathway [146, 170, 208]. This annotation covers 325 nt at ChrVIII:554148-554473 and was named *IMD2* upstream CUT.

3.5. GOStat analysis

GOStat analysis [209] was performed for all significantly down-regulated ORF-T transcripts from our dataset. In brief, the list of 995 significantly up-regulated ORF-Ts was entered into the GOStat web interface (http://gostat.wehi.edu.au/cgi-bin/goStat.pl) to search for the top 30 most over-represented GO terms and to obtain p-values to indicate the significance of enrichment (Table S3). Specifically, the GOStat tool counts the ratio at which each GO term appears in the test group and compares the counts to the reference gene list. A p-value is calculated for each GO term representing the probability that the term is enriched in the test group. A X^2-test is used to calculate the p-value unless the value for a count was below 5, in which case a

Fisher's exact test is more accurate [209]. GO-term enrichment analysis was also performed using DAVID (Database for Annotation Visualization and Integrated Discovery) [210]. DAVID is an integrated genomics data mining database that cross-references information from a number of online resources such as Gene Ontology (GO), NCBI, and Uniprot to extract a list of enriched biological processes. Similar results were obtained through GOStat and DAVID.

3.6. Northern Blot Analysis

30µg of total RNA was loaded per lane on a 1.5% agarose TBE gel and separated by electrophoresis at 120 volts for 1 hour at 4^{o}C. The gel was equilibrated in 10X SSC (Saline Sodium Citrate: 1.5M NaCl, 150mM Sodium Citrate) transfer buffer for 30 minutes. The RNA was transferred to Bio-Rad Zeta-Probe® blotting membranes by capillary overnight in 10X SSC. Transfer efficiency was determined by Methylene Blue staining. DNA oligonucleotide probes listed in Table 3 were 5' end-labeled with gamma ATP-^{32}P by T4 Polynucleotide Kinase. Probes were hybridized overnight to pre-blocked in Roche Life Science DIG Easy Hyb buffer at 37°C. Blots were washed in 6XSSC / 0.1%SDS once at room temperature and twice for 10 minutes at 50°C. Blots were exposed to a phosphorscreen overnight for snRNAs or 7 days for *FMP40* and *YPL22C-A* followed by scanning using a phosphorimager (GE Healthcare).

Name	Sequence	Reference
snR13 Probe	GCC AAA CAG CAA CTC GAG CCA AAT GCA CTC	This study
snR3 Probe	GCT CGA TCT TCG TAC TGT CTA ATG CGG TGG	This study
snR11 Probe	CTA TCA ACC GCG AGC ACG ACA GTG	This study
FMP40 Probe	GTA CCC AAC TTC TGG GGA CAA ACA ACG GG	This study
YPL222C-A Probe	CCC GTT GTT TGT CCC CAG AAG TTG GGT AC	This study

Table 3: Sequences for DNA oligonucleotide probes used for northern blotting.

4. ChIP-exo analysis of Rpb3-FLAG and Nrd1-TAP localization

4.1. Sample Preparation

Lysate preparation

Chromatin IP followed by exonuclease treatment was performed using the protocol described by Rhee and Pugh [211]. Rpb3-FLAG WT, *rtr1Δ* and *rrp6Δ* strains and Nrd1-TAP WT and *rtr1Δ* were grown to an OD_{600}=0.8-1 at 30^{o}C prior to crosslinking with formaldehyde (1% final concentration) for 15 minutes at 25^{o}C with shaking (Sigma, catalog # F8775-25ML). Crosslinking was immediately quenched with glycine (0.15M final concentration) for 5 minutes at 25^{o}C with shaking. Cells were pelleted by centrifugation for 3 minutes at 2,829 x g at 4^{o}C. Pellets were washed in 1ml ice-cold ST buffer (10mM Tris-HCl (pH 7.5),100mM NaCl) containing protease inhibitors, transferred to a 2ml Natural Conical tube, and pelleted at 2,300 x g for 2 min at 4 ^{o}C. Cells were frozen in -80 overnight to facilitate efficient lysis. Cells were thawed on ice and resuspended in 1ml ice-cold FA-lysis buffer (50 mM HEPES, 150 mM NaCl, 2 mM EDTA (pH 8.0), 1% Triton X-100, 0.1% Na Deoxycholate) containing protease inhibitors. One milliliter of 0.5-mm Zirconium Silicate beads were added to the suspension to facilitate yeast cell lysis. Cells were lysed in a bead-beater by mechanical lysis for 3 minutes, chilling on ice for 3 minutes, and repeating 3 times (9 minutes of lysis total). The conical tube was then punctured at the bottom with a red hot 22 gauge needle, placed inside a 1.7ml microfuge tube, and centrifuged at 200 x g briefly at room temperature to remove lysate from beads.

Chromatin Shearing by Sonication

Cells were centrifuged 2,300 x g for 3 minutes at 4^{o}C. The supernatant was discarded, and the cells were washed twice in 1ml ice-cold FA-lysis buffer containing protease inhibitors, centrifuged at 2,300 x g for 2 minutes at 4^{o}C each time. The pellet was then resuspended in 1 ml of ice-cold FA-lysis buffer containing 0.2% SDS and protease inhibitors and transferred to a TPX® Polymethylpentene (PMP) 15ml conical tube. Chromatin was sheared by sonication in a Diagenode Bioruptor in a 4^{o}C water bath on high for 30 cycles (30 seconds on / 30 seconds off). Lysate was transferred to a 1.7ml microfuge tube and centrifuged at 2,300 x g for 2 minutes at 4^{o}C to pellet an insoluble cellular debris. Supernatant was transferred back to the PMP tube, and sonication was repeated. Lysate was transferred to a fresh 1.5-ml microfuge tube and centrifuged at 2,300 × g for 10 minutes at 4°C, to pellet any remaining debris.

Chromatin Immunoprecipitation

The resulting supernatant from the sonication steps was transferred to a polypropylene 15ml conical tube, and 3ml FA-lysis buffer was added to dilute the SDS concentration to 0.05%. Immunoprecipitation was performed with 50uL of anti-FLAG agarose beads (for Rpb3-FLAG) or anti-TAP sepharose beads (for Nrd1-TAP) (Sigma). The volume of anti-FLAG agarose used for RNAPII immunoprecipitation was optimized by affinity purification followed by mass spectrometry. Chromatin and beads were incubated overnight with rotation at 4°C. The samples were centrifuged at 94 × g for 1 minute at room temperature to

collect the beads. The supernatant was removed, and beads were resuspended in 0.5ml ice-cold FA-lysis buffer containing protease inhibitors and transferred to a fresh 1.7ml microfuge tube. Beads were centrifuged at 94 x g for 1 minute at room temperature, the supernatant was removed, and the beads were washed in 1ml ice-cold FA-lysis buffer containing protease inhibitors by incubating 5 minutes at room temperature on a rotating wheel. Beads were pelleted by centrifugation at 94 × g for 1 minute at room temperature, resuspended in fresh ice-cold FA-lysis buffer containing protease inhibitors, and pelleted again to wash. The beads were then washed in the same manner in the following wash sequence, each buffer ice-cold and containing protease inhibitors:

1) 1ml FA-high salt wash buffer: 50mM HEPES, 1M NaCl, 2mM EDTA (pH 8.0), 1% Triton X-100, 0.1% NaDeoxycholate
2) 1ml FA-wash 2 buffer: 50mM Hepes/KOH, 0.5M NaCl, 2mM EDTA (pH 8.0), 1% TritonX-100, 0.1% NaDeoxycholate
3) 1ml FA-wash 3 buffer: 25mM LiCl, 1% NP40-Nonidet (IPEGAL), 1% NaDeoxycholate, 2mM EDTA, 10mM Tris-Cl (pH 8.0)
4) Tris-EDTA (TE) buffer: 10mM Tris-HCl (pH 8.0), 1mM EDTA (pH 8.0)
5) Tris-HCl (pH 8.0)

Each enzymatic reaction listed below was followed by the above wash sequence.

On-Bead Enzymatic Reactions – First Adaptor Ligation and Exonulcease Digestion for ChIP-exo

A diagram of these enzymatic reactions is shown in Figure 8.

1) The DNA bound to the immunoprecipitated protein was subjected to “polishing” by T4 DNA polymerase (New England Biolabs). This reaction served to fill in DNA sequences in the 5’ – 3’ direction at the end of DNA fragments left with a 5’ single stranded overhang after sonication, resulting in blunt ends.

2) Next, T4 Polynucleotide Kinase (PNK) (New England Biolabs) was used to add a phosphate to the 5’ –hydroxyl ends of the blunt ends. This step increases the efficiency of the ligation reaction used to add barcoded adaptors to the DNA for sequencing.

3) Klenow Fragment (New England Biolabs) and dATP were used to add a polyA tail to the DNA 3’ ends. This polyA tail is used for ligation of the P2-T adaptors.

4) T4 DNA ligase (New England Biolabs) was used to join a P2-T barcoded sequencing adaptor to the blunt ends of the DNA fragments. Samples and their corresponding barcodes are listed in Table 5.

5) Filling in reactions were then performed with Phi29 DNA polymerase (New England Biolabs) – a replicative polymerase – to fill in the adaptor sequence.

6) A second T4 PNK (New England Biolabs) reaction added a phosphate to the 5’ –hydroxyl ends to facilitate the exonuclease reaction.

7) Lambda exonuclease (New England Biolabs) was used to digest the 5’ strand of the double stranded DNA back to the location of the bound immunoprecipitated protein. Lambda exonuclease is capable of degrading single stranded DNA and unphosphorylated ends but is much more efficient digesting

double stranded, 5’ –phosphorylated ends according to the manufacturer’s instructions.

8) A second 5’ – 3’ exonuclease reaction was performed with $RecJ_f$ exonuclease (New England Biolabs) to maximize the efficiency of the exonuclease digestion.

Chromatin Elution

Chromatin was eluted from the beads and the cross-linking was reversed by incubating in 450µl ChIP Elution Buffer (25mM Trizma 2mM EDTA (pH 8.0), 200mM NaCl, 0.5% SDS) with 1µl Protease K at 65°C overnight. The DNA was isolated with ice-cold phenol/chloroform/isoamyl alcohol (Ambion) and precipitated for 1 hour with 100% ethanol. The pellets were dried, resuspended in 11µl TE, and immediately sent to the sequencing core for the addition of the second sequencing adaptor, DNA quality and size confirmation by BioAnalyzer, and SOLiD5500xl sequencing. BioAnalyzer runs were performed according to the manufacturer’s instructions on a high sensitivity DNA chip.

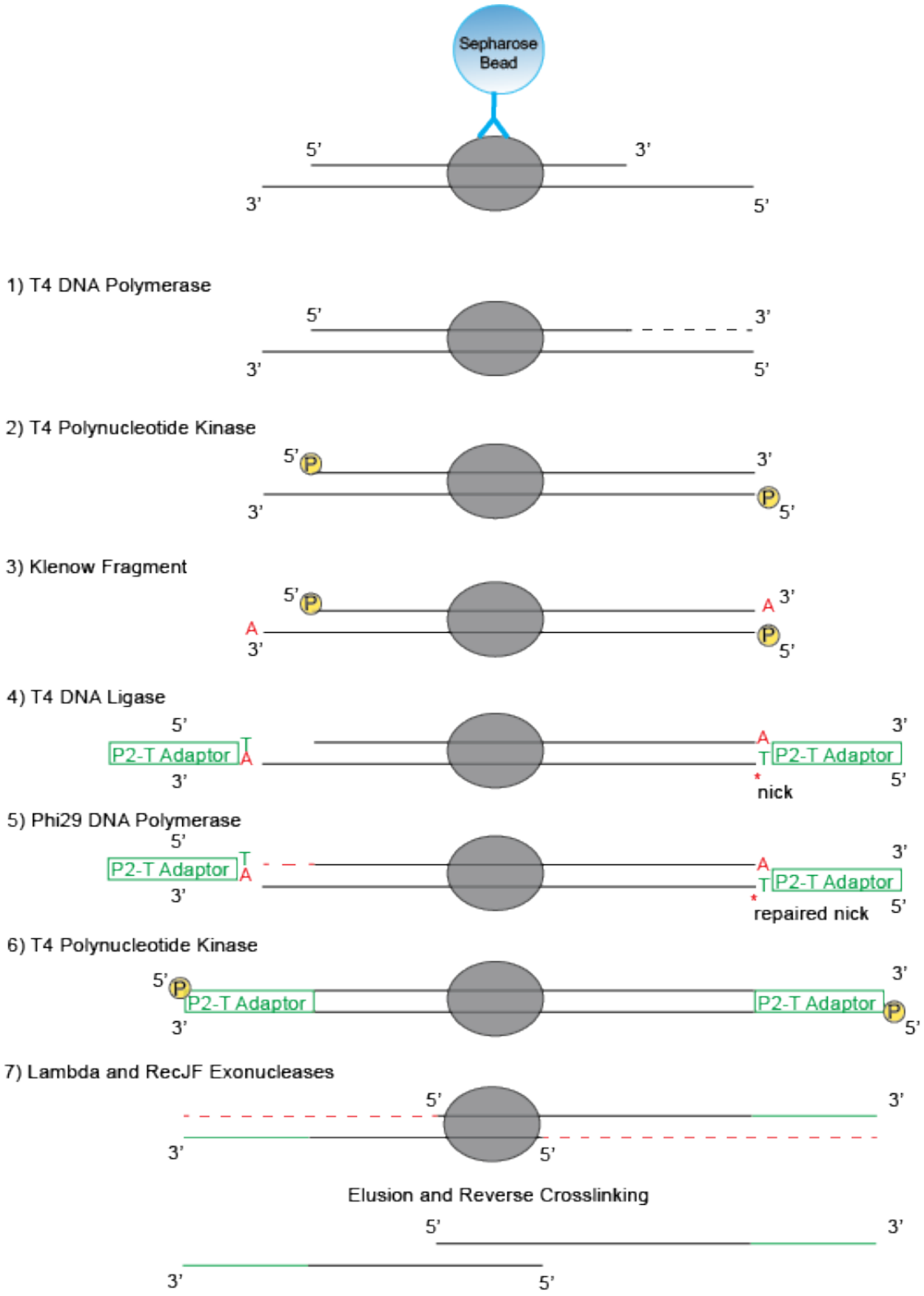

Figure 8: Schematic of enzymatic reactions for ChIP-exo sample preparation. Reactions include adaptor ligation and exonuclease digestion of 5' ends.

4.2. ChIP-exo peak calling with MACS2

To identify the top peaks of Nrd1-RNAPII binding in the Nrd1 ChIP-exo data set, we used the MACS2 (Model-based Analysis of ChIP-Seq) algorithm [212]. While the original MACS protocol was developed for calling narrow transcription factor binding sites [213], MACS2 is capable of calling broad peaks similar to the footprint expected of Nrd1 binding. We executed MACS2 (Version 2-2.1.0) for Nrd1-TAP WT (Library 29) and Nrd1-TAP *rtr1Δ* cells (Library 30) independently. Default parameters were used with the fragment extension length set to 300 bases. This length was determined by visual evaluation of the mapped DNA reads in IGV. The genomic locations of the top 100 peaks as determined by fold-change were compared to the genomic locations of the top 100 Nrd1 Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) peaks reported by Creamer *et al.* [50].

RESULTS

1. Confirmation of strains and sample quality

1.1. Transformation of knockout strains

All yeast strains used are isogenic to BY4741, which was used as our wild-type in all experiments. The *RRP6* deletion strain was obtained from the yeast knockout collection (Open Biosystems) [198]. *RTR1* was knocked out of BY4741 and the *RRP6* deletion strain by homologous recombination by using a kanamycin cassette (containing the *KANMX* gene) to create the *RTR1* deletion and *RTR1/RRP6* double deletion strains (illustrated in Figure 9). The *RTR1::KANMX* strain was previously constructed in the lab, so to knock out *RTR1* in the *rrp6Δ* cells, the *RTR1::KANMX* locus plus 200 base pairs upstream of the start codon and 200 base pairs downstream of the stop codon were amplified by polymerase chain reaction using primers P1 and P2 (Figure 9). The resulting PCR product was then transformed into the *rrp6Δ* cells to stably delete the *RTR1* locus by homologous recombination and transformants were then selected for kanamycin resistance. *RTR1* is 681 nt and *KANMX* is 1400 nt. Confirmation of successful transformation was performed by PCR using primers P3 and P4 (Figure 9). Successful recombination of the transformed DNA resulted in a PCR product of 2147 nt whereas an unsuccessful transformation would result in a PCR product of 1428 nt. Representative PCR results are shown in Figure 10.

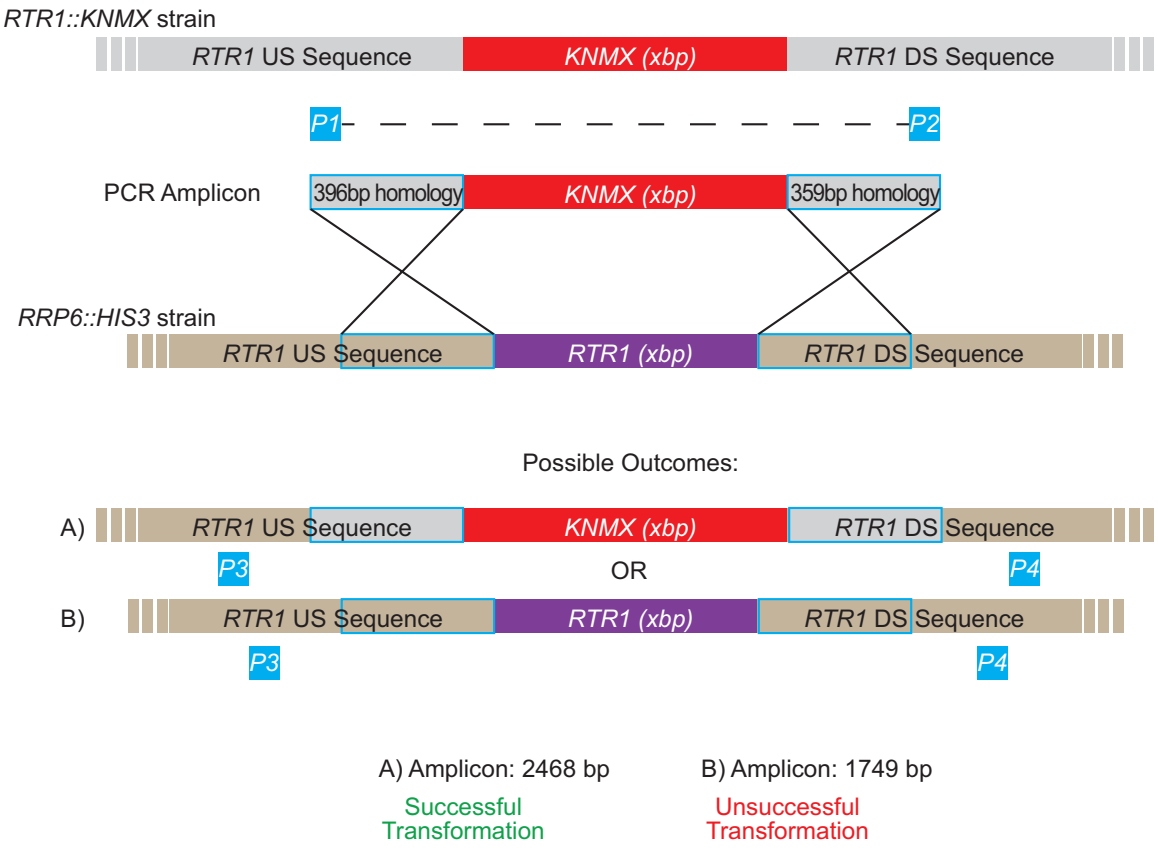

Figure 9: Schematic of homologous recombination and PCR confirmation. The *KNMX* sequence and 396 nt homologous to the sequence upstream of *RTR1* and 359 nt homologous to the sequence downstream of *RTR1* were amplified by PCR using primers P1 and P2 from the available *RTR1::KNMX* strain. The resulting PCR amplicon was used to transform the *RRP6::HIS3* strain that was obtained from the Open Biosystems yeast knockout library by homologous recombination. Successful transformation was confirmed by PCR using primers P3 and P4.

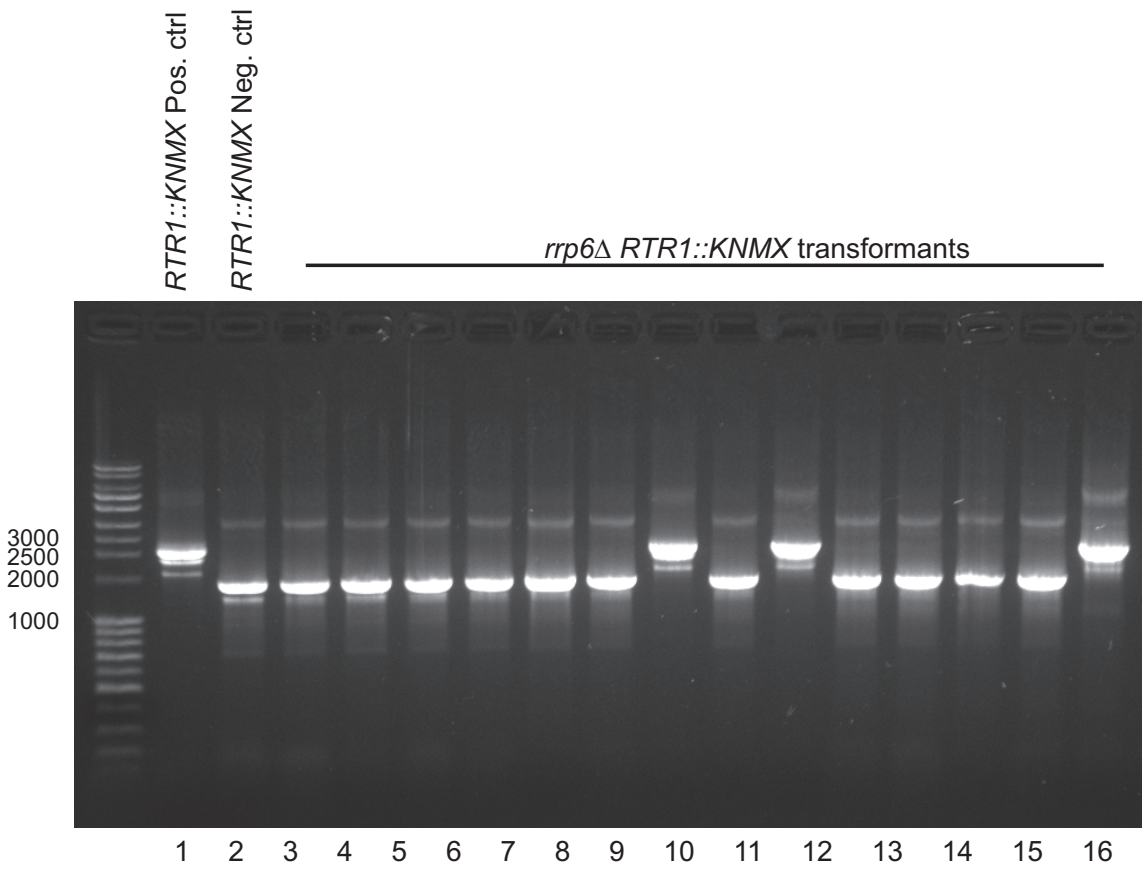

Figure 10: Confirmation of *RTR1* knockout by PCR. PCR products resulting from amplification of the *RTR1* region using primers P4 and P5 (Figure 9). A positive knockout control where *RTR1* has been replaced with *KNMX* is in lane 1 at 2468 bp. A negative control where *RTR1* is present is in lane 2 at 1749 bp. Results for 14 transformants tested are in lanes 3-16. Transformants with the desired *RTR1::KNMX* recombination can be seen in lanes 10, 12, and 16. All three positives were frozen for future use and tested with comparable results in early experiments. All three performed identically in all experiments. The strain represented in lane 10 was used for all analysis shown here.

1.2. Sample quality confirmation by Bioanalyzer

For RNA isolation, nine biological replicates were streaked from glycerol stocks for each of four strains: WT, *rtr1Δ*, *rrp6Δ*, and *rtr1Δ rrp6Δ* . Each sample set was confirmed by electrophoresis and PCR amplification to be of good quality before sent for analysis. For ChIP-exo preparation, four replicates were performed for Rpb3-FLAG WT and *rtr1Δ*, two replicates were performed for Rpb3-FLAG *rrp6Δ*, and three replicates were performed for Nrd1-TAP WT and *rtr1Δ*. RNA and DNA quality were ascertained by BioAnalyzer. The Agilent Bioanalyzer uses a glass chip with multiple microchannels. The samples and a known concentration of a ladder are loaded into each well with fluorescent dye. A voltage gradient is used to separate the RNAs and DNA fragments by size, similar to gel electrophoresis. Each RNA and DNA molecule is detected by laser-induced fluorescence, and the data is translated into gel-like images and peaks. The software compares the unknown peaks detected in the RNA and DNA samples to the known sizes and concentration of the bands detected in the ladder to calculate the concentration of the samples. In the case of the RNA, the expected sizes and concentrations of the yeast ribosomal RNA bands, as well as other highly abundant bands, are compared to the detected sizes and concentrations to calculate the RNA Integrity Number (RIN) representing the quality of the RNA sample (Table 4). RINs of 6 or greater are considered acceptable for RNA-Seq. The best four replicates as determined by Bioanalyzer were selected for sequencing: P5, P6, P8, and P9 (Figure 11, Table 4). For ChIP-exo samples, the resulting peaks were compared to the known peaks of the ladder to calculate the average fragment size (Figure

12, Table 5). An average DNA fragment size of approximately 250 nt is optimal for ChIP-exo experiments [211].

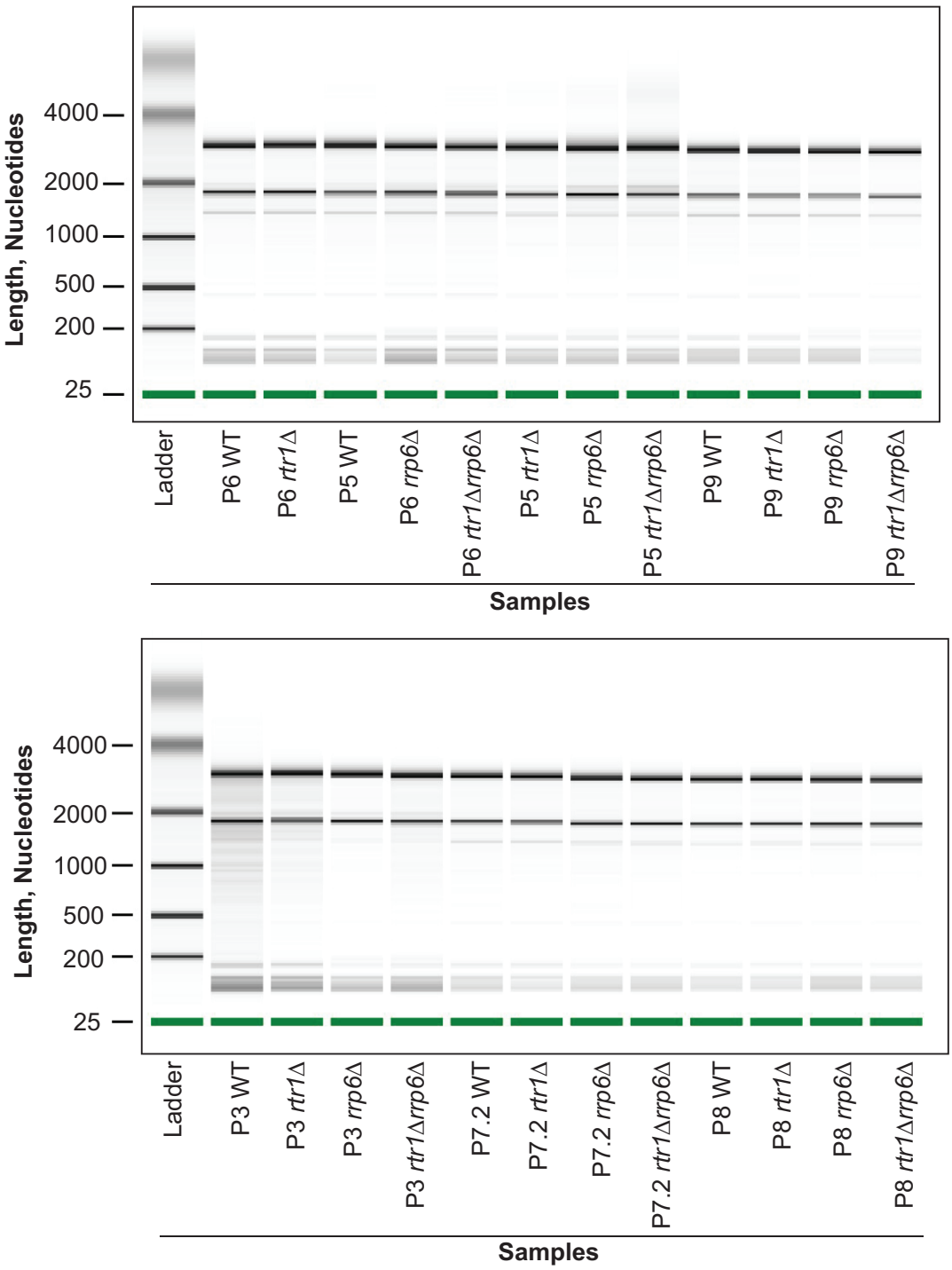

Figure 11: RNA Bioanalyzer results. The top and bottom panels represent the results from one chip. Each RNA preparation is listed as P# (i.e. P3) corresponding to the biological replicate purification and was used for further analysis. The RNA ladder is in the left-most lane of each chip. The green bar indicates a small RNA control marker added to each lane to serve as a loading control.

Sample	Barcode ID	RIN
WT prep 8	1	6.50
rtr1Δ prep 8	2	6.60
rrp6Δ prep 8	3	6.40
rtr1Δ rrp6Δ prep 8	4	6.50
WT prep 5	5	6.50
rtr1Δ prep 5	6	6.70
rrp6Δ prep 5	7	6.50
rtr1Δ rrp6Δ prep 5	8	8.80
WT prep 6	9	6.40
rtr1Δ prep 6	10	6.40
rrp6Δ prep 6	11	6.40
rtr1Δ rrp6Δ prep 6	12	6.30
WT prep 9	13	6.50
rtr1Δ prep 9	14	6.60
rrp6Δ prep 9	15	6.70
rtr1Δ rrp6Δ prep 9	16	6.60

Table 4: Summary of Bioanalyzer data for RNA samples used in this study. RIN (RNA Integrity Number) values were calculated by the Bioanalyzer software which compares the peaks shown in Figure 11 to the known sizes of the *S. cerevisiae* rRNAs. RINs of 6 or greater are considered acceptable for RNA-Seq.

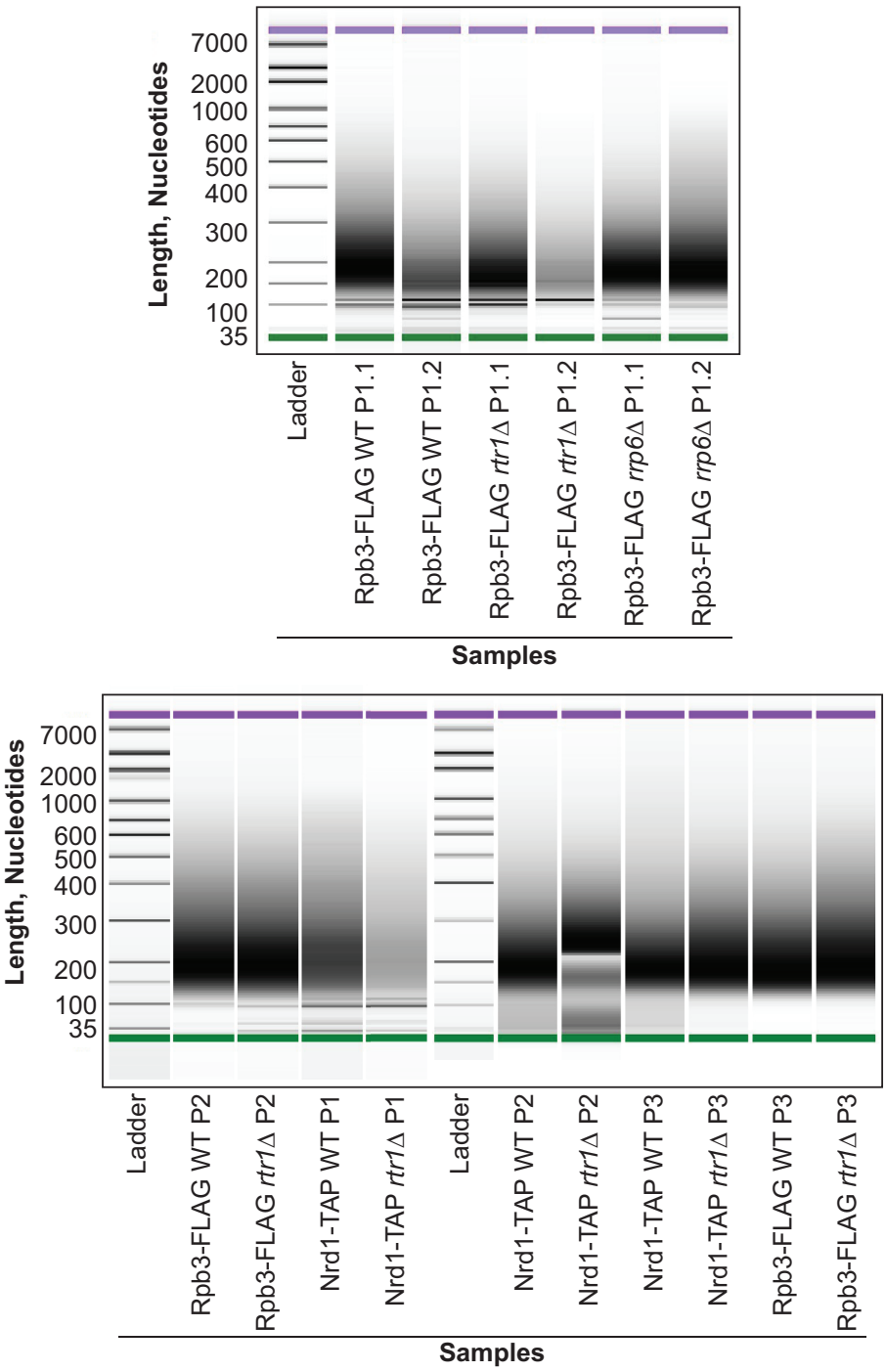

Figure 12: ChIP-exo Bioanalyzer results. The top panel represents the results from one chip. The bottom panel represents two partial chips combined for space. The DNA ladder is in the left-most lane of each chip. Sheared DNA samples are shown in the other lanes. The green and purple bars indicate loading control markers added to each lane.

Sample	Barcode ID	Ave Fragment Size, nt
Rpb3-FLAG WT p1	21	243
Rpb3-FLAG *rtr1Δ* p1	23	225
Rpb3-FLAG *rrp6Δ* p1	25	229
Rpb3-FLAG WT p2	28	264
Rpb3-FLAG *rtr1Δ* p2	27	260
Nrd1-TAP WT p1	29	285
Nrd1-TAP *rtr1Δ* p1	30	259

Table 5: Summary of ChIP-exo Bioanalyzer results for samples used in this study. Average DNA fragment sizes were calculated by using the Bioanalyzer software which compares the peaks for each sample in Figure 12 to the size and intensity of the DNA ladder. An average DNA fragment size of approximately 250 nt is optimal for ChIP-exo experiments [211]. The p# designates the biological replicate number for the samples used for analysis.

2. Rrp6 Is Required for RNAPII Termination at Specific Targets of the Nrd1-Nab3 Pathway

2.1. Genome-wide analysis of *RRP6* deletion strains by RNA-Seq

To identify classes of transcripts affected by the loss of Rrp6, we utilized previously published annotations for yeast transcripts and performed differential expression analysis using normalized sequencing reads through EdgeR [72, 205, 214]. This analysis represented the entire transcribed region for mRNAs by employing annotations that include both the 5' and 3' untranslated region (UTR) for the majority of the yeast transcriptome [215]. Overall, the annotations used for our study include 5792 open reading frames (ORF-Ts), 658 CUTs, 648 SUTs, 1215 NUTs, 844 Rrp6-regulated antisense transcripts, 1078 Xrn1-sensitive unstable transcripts (XUTs), 80 snRNAs, and 78 snRNA extended transcripts (ETs, manually annotated in this study). In *rrp6Δ* cells, we identified 136 up-regulated open reading frame transcripts and 734 that were down-regulated (ORF-Ts, fold change cut-off = +/- 1.5, p-value < 0.05, FDR ≤ 0.1, Figure 13A, B). We also identified 622 induced CUTs out of a total of 733, in agreement with previous reports that CUT expression and stability is increased in the absence of Rrp6 ([71, 73, 75], Figure 13A). Of importance, using deep sequencing technology many CUTs were detected that have not previously been identified in WT cells using microarrays [71-73, 75]. Analysis of the NUTs revealed that the majority of this class of transcripts, 887/1215, were also up-regulated in the *rrp6Δ* strain (Figure 14A). We also discovered that 223/648 SUTs were significantly up-regulated in *rrp6Δ* cells, identifying the specific subset of SUT transcripts that are

likely terminated by the NNS pathway. In *rrp6Δ* cells, a total of 54 sn/snoRNAs showed significant transcript extension. There was not significant overlap between Rrp6-sensitive transcripts and previously described XUTs, in agreement with previous work that suggests that noncoding RNAs targeted by the cytoplasmic 5'-3' exonuclease Xrn1 are degraded by a separate mechanism than noncoding RNAs sensitive to the nuclear specific 3'-5' exonuclease Rrp6 [216].

Considering that Rrp6 is an exonuclease, it is expected that the majority of RNA expression changes in *rrp6Δ* cells would be due to accumulation. However, we were surprised to see 734 mRNAs significantly decreased ≤ 1.5-fold in *rrp6Δ* cells (Figure 13B, ORF-Ts in yellow). To better understand the classes of mRNAs down-regulated in *rrp6Δ* cells, we performed GO-term enrichment analysis to determine if any cellular pathways showed significant enrichment within the set of decreased mRNAs. Surprisingly, the most enriched GOterm was GO:0005830 for the cytosolic ribosome with a p-value of $6.43E^{-82}$ (Table 6). In total, nine of the 30 statistically significant enriched GO-terms related to ribosomal protein coding genes with p-values less than or equal to $7.85E^{-19}$ (Table 6). Differential expression analysis determined that transcript levels of 113 out of 137 ribosomal protein coding genes were decreased more than 1.5-fold (Figure 13C, Table 3). Interestingly, only three ribosomal protein-coding genes had more than 1.5-fold increase in expression (Figure 13C). It is also interesting to note that the average transcript length from the down-regulated ribosomal protein-coding mRNAs is 916 nt putting many of the transcripts within the approximate length limits for the NNS pathway.

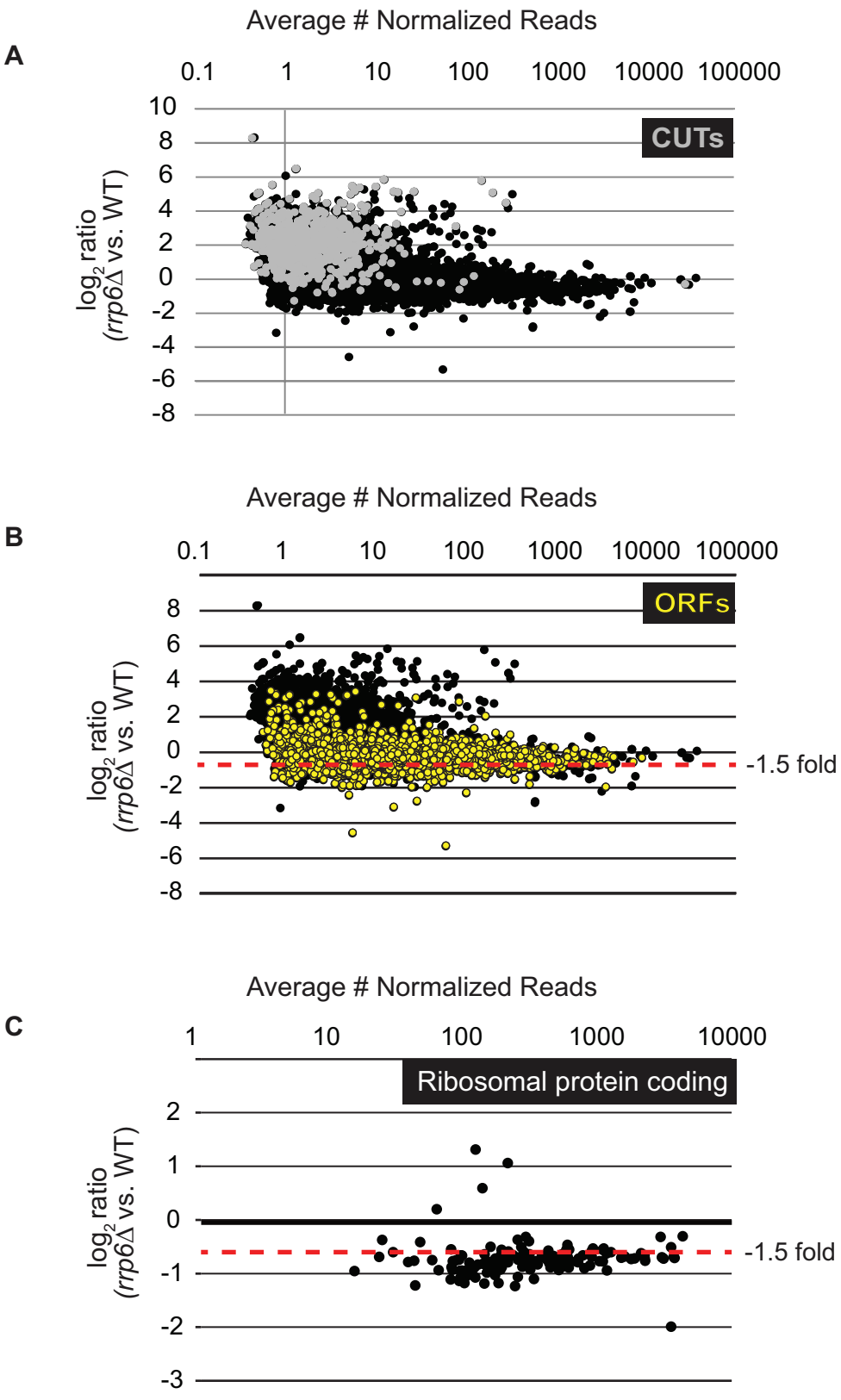
A
Average # Normalized Reads
0.1
1
10
100
1000
10000
100000
10
8
6
4
2
0
-2
-4
-6
-8
log2 ratio
(rrp6Δ vs. WT)
CUTs
B
Average # Normalized Reads
0.1
1
10
100
1000
10000
100000
8
6
4
2
0
-2
-4
-6
-8
log2 ratio
(rrp6Δ vs. WT)
ORFs
-1.5 fold
C
Average # Normalized Reads
1
10
100
1000
10000
Ribosomal protein coding
2
1
0
-1
-2
-3
log2 ratio
(rrp6Δ vs. WT)
-1.5 fold

Figure 13: Expression plots for normalized RNA-Seq data with specific classes of RNA transcripts highlighted. After sequencing reads were aligned to the yeast genome, reads mapped to annotated open reading frame transcripts (ORF-Ts), cryptic unstable transcripts (CUTs), stable unannotated transcripts (SUTs), and Nrd1 unterminated transcripts (NUTs) were used for differential expression analysis using edgeR. Log_2 of the fold-change values are plotted versus the average number of normalized reads across all biological replicates for all RNA transcripts in cells lacking *RRP6* compared to WT (black dots). (A) RNAs previously annotated as CUTs, a classification based on the dependence of *rrp6Δ* for detection are shown as gray dots while all other transcripts are shown as black dots. (B) RNAs annotated as ORF-Ts, most of which are protein coding messenger RNAs are shown in yellow. All other transcripts are shown in black. (C) Messenger RNA expression values for ribosomal protein coding genes, shown as black dots.

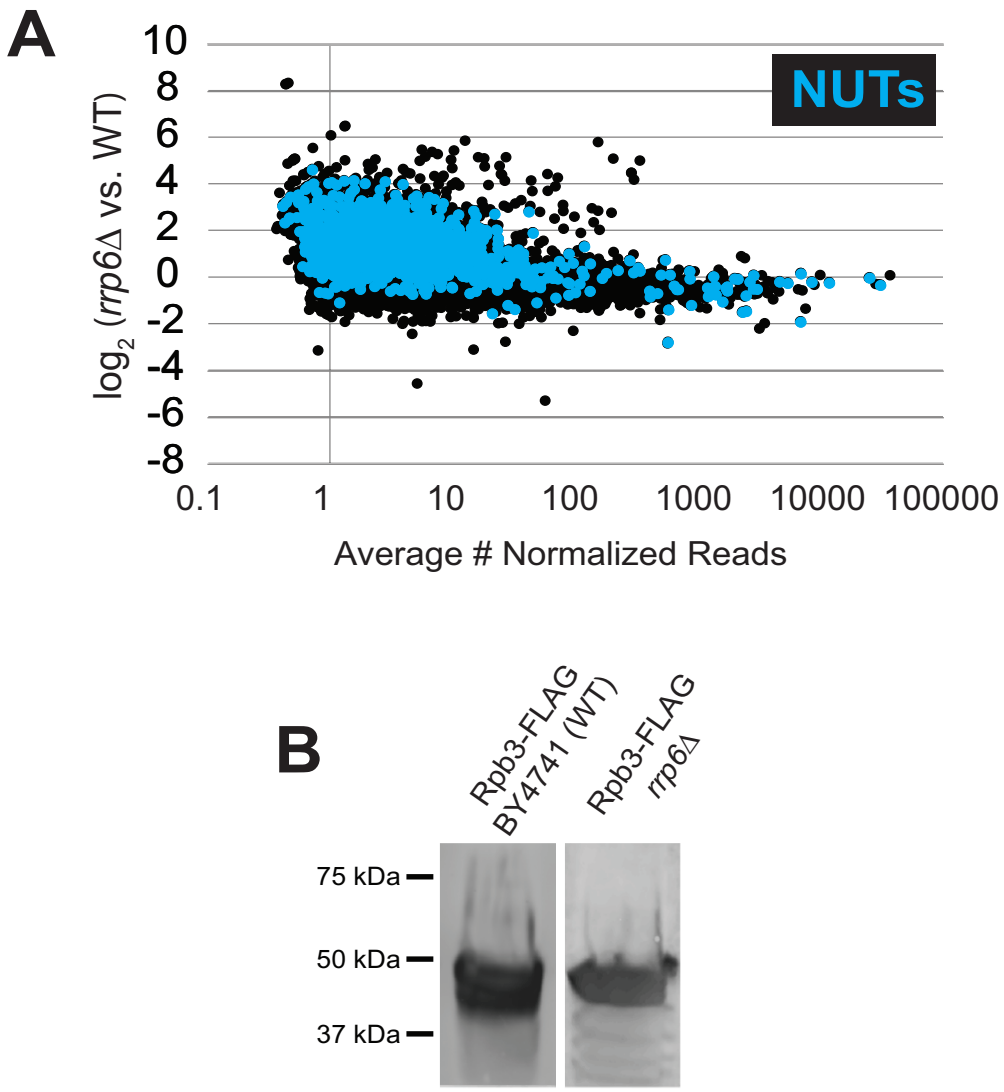

Figure 14: Expression plots for normalized RNA-Seq data for Nrd1-unterminated transcripts (NUTs) and western blot analysis of Rpb3-FLAG strains. (A) After sequencing reads were aligned to the yeast genome, reads mapped to annotated open reading frame transcripts (ORF-Ts), cryptic unstable transcripts (CUTs), stable unannotated transcripts (SUTs), and Nrd1 unterminated transcripts (NUTs) were used for differential expression analysis in *rrp6Δ* versus WT using edgeR. Log_2 of the fold-change values are plotted versus the average number of normalized reads across all biological replicates for all RNA transcripts in cells lacking *RRP6* compared to WT (black dots). NUT annotations are shown as aqua dots. (B) Western blot analysis of whole cell extracts prepared from Rpb3-FLAG WT and *rrp6Δ* strains using anti-FLAG peroxidase coupled antibodies (Sigma).

Best GOs (Max: 30)	Count 927	Total 6476	P-Value	Ontology
GO:0005830	115	175	6.43E-82	Cytosolic Ribosome
GO:0044445	120	196	6.71E-77	Cytosolic Part (Cytosol Component)
GO:0003735	122	230	9.32E-62	Structural Constituent of Ribosome
GO:0033279	123	240	8.96E-59	Ribosome
GO:0005842	68	97	1.37E-52	Cytosolic Large Ribosomal Subunit
GO:0005840	139	353	2.12E-40	Ribosome
GO:0015934	76	142	3.29E-38	Large Ribosomal Subunit
GO:0005843	45	64	2.49E-34	Cytosolic Small Ribosomal Subunit
GO:0043228	259	1032	1.65E-24	Non-Membrane-Bound Organelle
GO:0043232	259	1032	1.65E-24	Intracellular Non-Membrane-Bound Organelle
GO:0030529	177	622	1.38E-23	Ribonucleoprotein Complex
GO:0015935	47	98	7.85E-19	Small Ribosomal Subunit
GO:0009277	46	114	5.68E-13	Fungal-Type Cell Wall
GO:0005618	46	114	5.68E-13	Cell Wall
GO:0030312	46	114	5.68E-13	External Encapsulating Structure
GO:0005829	145	623	4.07E-09	Cytosol
GO:0044422	420	2309	5.66E-09	Organelle Part
GO:0044446	420	2309	5.66E-09	Intracellular Organelle Part
GO:0043229	666	4023	6.50E-09	Intracellular Organelle
GO:0043226	666	4024	6.63E-09	Organelle
GO:0044249	209	1009	3.93E-08	Cellular Biosynthetic Process
GO:0032991	322	1705	4.54E-08	Macromolecular Complex
GO:0044464	846	5490	3.67E-07	Cell Part
GO:0006412	145	687	9.57E-06	Translation
GO:0005737	602	3699	2.15E-05	Cytoplasm
GO:0009058	236	1251	3.37E-05	Biosynthetic Process
GO:0044444	458	2717	6.65E-05	Cytoplasmic Part
GO:0009987	729	4653	6.65E-05	Cellular Process
GO:0005199	12	19	0.000102	Structural Constituent of Cell Wall
GO:0030150	13	22	0.000104	Protein Import into Mitochondrial Matrix

Table 6: Top 30 GO-terms enriched in protein coding genes down-regulated in *rrp6Δ* compared to WT. Table includes GO identification number (“Best GOs”), number of hits from GOStat analysis of our list of protein coding genes down-regulated in *rrp6Δ* matching the GO ID (“Count”), total number of genes assigned the GO ID (“Total”), p-value for the GO term, and the descriptor for the GO ID. Information calculated using GOstat, as described in the methods.

Transcript name	Standard gene name	RRP6 _KO vs WT log2_FC	RRP6 _KO vs WT p-value	RRP6 _KO vs_WT FDR
YLR367W	*RPS22B*	-1.992743	1.27E-06	1.34E-05
YGL189C	*RPS26A*	-1.234646	4.08E-04	2.11E-03
YDR012W	*RPL4B*	-1.224361	2.50E-05	1.80E-04
YOL040C	*RPS15*	-1.187806	1.48E-04	0.000871
YOL039W	*RPP2A*	-1.18084	2.86E-04	1.54E-03
YKL180W	*RPL17A*	-1.178588	2.47E-04	1.36E-03
YLR185W	*RPL37A*	-1.138591	7.94E-05	0.000498
YDR418W	*RPL12B*	-1.112196	1.63E-05	0.000123
YBL092W	*RPL32*	-1.103358	0.0002222	0.001236
YHL033C	*RPL8A*	-1.096525	2.78E-05	0.000197
YER131W	*RPS26B*	-1.075197	1.52E-04	0.00089
YCR031C	*RPS14A*	-1.067605	0.0002575	0.001406
YOL127W	*RPL25*	-1.061722	0.0006462	0.00316
YMR242C	*RPL20A*	-1.035553	0.0010762	0.005013
YML073C	*RPL6A*	-1.032985	0.000519	0.002613
YPL249C-A	*RPL36B*	-0.983926	0.0006543	0.003197
YPL079W	*RPL21B*	-0.964466	0.0001339	0.000795
YFL034C-A	*RPL22B*	-0.955394	0.0012336	0.005662
YDR025W	*RPS11A*	-0.949811	0.0016071	0.007215
YIL052C	*RPL34B*	-0.948519	0.0009384	0.004451
YOL120C	*RPL18A*	-0.943612	0.0004191	0.002161
YLL045C	*RPL8B*	-0.943453	0.000723	0.00351
YNL096C	*RPS7B*	-0.938209	0.0005994	0.002957
YNL302C	*RPS19B*	-0.93574	0.0013151	0.006009
YER056C-A	*RPL34A*	-0.930519	0.0009908	0.004656
YER102W	*RPS8B*	-0.928462	6.59E-04	3.22E-03
YKL006W	*RPL14A*	-0.92515	1.58E-03	7.09E-03
YNL178W	*RPS3*	-0.901668	0.0025041	0.010605
YBL087C	*RPL23A*	-0.892906	0.002011	0.008786
YGL031C	*RPL24A*	-0.88995	0.0016174	0.007259
YLR048W	*RPS0B*	-0.885436	0.0023533	0.010064
YIL133C	*RPL16A*	-0.885136	0.0004631	0.002356
YPR043W	*RPL43A*	-0.884017	0.002664	0.011217
YOL121C	*RPS19A*	-0.883042	0.0010284	0.004808
YER074W	*RPS24A*	-0.87574	0.0058453	0.022194
YBR189W	*RPS9B*	-0.869864	0.0024119	0.010275
YDR500C	*RPL37B*	-0.866716	0.0017226	0.007679
YFR031C-A	*RPL2A*	-0.851989	0.005231	0.02018
YPL131W	*RPL5*	-0.843837	0.0045064	0.017737
YOR182C	*RPS30B*	-0.835264	0.0078704	0.028345
YDL083C	*RPS16B*	-0.834289	0.0044297	0.017518

Transcript name	Standard Gene Name	RRP6 _KO vs WT log2_FC	RRP6 _KO vs WT p-value	RRP6 _KO vs_WT FDR
YBL072C	*RPS8A*	-0.831343	0.0019911	0.008726
YLR333C	*RPS25B*	-0.819168	0.0034659	0.014144
YBR031W	*RPL4A*	-0.808244	0.0048695	0.018981
YOR312C	*RPL20B*	-0.806282	0.006965	0.025671
YLR061W	*RPL22A*	-0.803565	0.0086104	0.030498
YDL061C	*RPS29B*	-0.798657	0.0044449	0.01755
YLR448W	*RPL6B*	-0.798618	0.0047792	0.018658
YDR064W	*RPS13*	-0.791903	0.0052173	0.020134
YNL069C	*RPL16B*	-0.789254	0.0137802	0.045367
YIL069C	*RPS24B*	-0.787525	0.0060321	0.02279
YJL177W	*RPL17B*	-0.786743	0.0076389	0.027709
YBL027W	*RPL19B*	-0.786163	0.0135939	0.044915
YER117W	*RPL23B*	-0.778262	0.0095956	0.03347
YGR027C	*RPS25A*	-0.774556	0.0114579	0.039038
YGL030W	*RPL30*	-0.772552	0.0080671	0.028908
YDL136W	*RPL35B*	-0.770191	0.0040663	0.016259
YLR340W	*RPP0*	-0.77013	0.0110787	0.037909
YOR096W	*RPS7A*	-0.764428	0.0204919	0.062694
YLR388W	*RPS29A*	-0.761878	0.0042301	0.016807
YLR325C	*RPL38*	-0.759297	0.0085087	0.030212
YDL075W	*RPL31A*	-0.755111	0.0065852	0.024522
YFR032C-A	*RPL29*	-0.754373	0.0444649	0.11698
YLR264W	*RPS28B*	-0.753698	0.0357407	0.098824
YDL130W	*RPP1B*	-0.751444	0.0019945	0.00873
YGR214W	*RPS0A*	-0.742393	0.010866	0.037321
YOR293W	*RPS10A*	-0.74227	0.014457	0.04721
YGR085C	*RPL11B*	-0.74147	0.0085608	0.030354
YLR029C	*RPL15A*	-0.731306	0.0227316	0.068083
YOR234C	*RPL33B*	-0.728371	0.0232336	0.069254
YDR450W	*RPS18A*	-0.727014	0.0146413	0.047688
YJR123W	*RPS5*	-0.725396	0.0133131	0.044146
YPL143W	*RPL33A*	-0.724666	0.0117044	0.0397
YMR142C	*RPL13B*	-0.724028	0.0118811	0.040184
YGL103W	*RPL28*	-0.721152	0.0154361	0.049858
YIL018W	*RPL2B*	-0.719007	0.0215832	0.065339
YHR203C	*RPS4B*	-0.717286	0.009648	0.03363
YKR094C	*RPL40B*	-0.713796	0.0177864	0.056102
YOR063W	*RPL3*	-0.711676	1.17E-02	3.96E-02
YGL147C	*RPL9A*	-0.708731	1.98E-02	6.10E-02
YNL162W	*RPL42A*	-0.701371	0.0140184	0.046031
YGL123W	*RPS2*	-0.699459	0.015224	0.049315

Transcript name	Standard gene name	RRP6 _KO vs WT log2_FC	RRP6 _KO vs WT p-value	RRP6 _KO vs_WT FDR
YIL148W	*RPL40A*	-0.69744	0.0256972	0.075447
YDL081C	*RPP1A*	-0.692989	0.0138284	0.0455
YPL090C	*RPS6A*	-0.692649	0.0128554	0.042953
YLR167W	*RPS31*	-0.692286	0.017123	0.054322
YMR194W	*RPL36A*	-0.690894	0.0145699	0.047502
YML024W	*RPS17A*	-0.685614	0.028975	0.083433
YLR441C	*RPS1A*	-0.682298	0.01404	0.046087
YHR141C	*RPL42B*	-0.675091	0.0171414	0.054357
YEL054C	*RPL12A*	-0.672709	0.0089589	0.031566
YML063W	*RPS1B*	-0.666632	0.0194106	0.060042
YLR406C	*RPL31B*	-0.666172	0.0144954	0.047289
YBR181C	*RPS6B*	-0.664573	0.0167511	0.053404
YDL191W	*RPL35A*	-0.662184	0.0153436	0.049623
YJL190C	*RPS22A*	-0.661559	0.0261498	0.076566
YHL015W	*RPS20*	-0.658424	0.019602	0.060523
YLR287C-A	*RPS30A*	-0.657464	0.017669	0.055767
YBR048W	*RPS11B*	-0.65009	0.0199817	0.061432
YDR471W	*RPL27B*	-0.644312	0.0258483	0.075794
YKR057W	*RPS21A*	-0.642271	0.0276957	0.080346
YGR148C	*RPL24B*	-0.642236	0.0182273	0.057082
YHL001W	*RPL14B*	-0.639806	0.0232776	0.069365
YDL082W	*RPL13A*	-0.634606	0.0175214	0.055387
YDR447C	*RPS17B*	-0.626113	0.029302	0.084242
YOR369C	*RPS12*	-0.624058	0.0206402	0.063052
YBR191W	*RPL21A*	-0.622635	0.0416636	0.111294
YGR034W	*RPL26B*	-0.61042	0.0178033	0.056138
YPR102C	*RPL11A*	-0.608627	0.0229386	0.068598
YLR344W	*RPL26A*	-0.603486	0.0192758	0.059735
YHR010W	*RPL27A*	-0.601308	0.0493884	0.126694
YMR143W	*RPS16A*	-0.599466	0.0346025	0.096177
YDR382W	*RPP2B*	-0.589555	0.0261842	0.0766
YPR132W	*RPS23B*	-0.583809	0.0406487	0.109395
YOR167C	*RPS28A*	-0.576323	0.0495282	0.126891
YHR021C	*RPS27B*	-0.560382	0.0628014	0.152233
YGR118W	*RPS23A*	-0.559219	0.042936	0.113849
YML026C	*RPS18B*	-0.557373	0.059988	0.146896
YGL135W	*RPL1B*	-0.55551	0.0363786	0.100071
YJL136C	*RPS21B*	-0.549716	0.0446804	0.117424
YJR094W-A	*RPL43B*	-0.536115	0.0754008	0.17443
YJR145C	*RPS4A*	-0.533175	0.0835771	0.188317
YLR075W	*RPL10*	-0.512194	0.0608524	0.148472

Transcript name	Standard gene name	RRP6 _KO vs WT log2_FC	RRP6 _KO vs WT p-value	RRP6 _KO vs_WT FDR
YJL189W	*RPL39*	-0.505149	0.0831211	0.18768
YMR230W	*RPS10B*	-0.503579	0.0968439	0.210189
YPL220W	*RPL1A*	-0.468115	0.0691459	0.163534
YKL156W	*RPS27A*	-0.41425	0.1248119	0.252818
YGL076C	*RPL7A*	-0.39988	0.1361771	0.270211
YHR062C	*RPP1*	-0.376622	0.1612654	0.303452
YNL067W	*RPL9B*	-0.371443	0.1648438	0.308062
YDL133C-A	*RPL41B*	-0.319501	0.2893062	0.453218
YBR084C-A	*RPL19A*	-0.317091	0.2983566	0.46308
YDL184C	*RPL41A*	-0.307528	0.3224259	0.486798
YPL198W	*RPL7B*	0.1986869	0.5286336	0.665783
YPL081W	*RPS9A*	0.5890903	0.0698796	0.164873
YNL301C	*RPL18B*	1.060323	4.82E-05	0.000323
YJL191W	*RPS14B*	1.3124904	2.56E-06	2.49E-05

Table 7: Differential expression information for the ribosomal protein coding transcripts. Table includes differential expression data expressed in $\log_2$ *rrp6Δ*/WT ratio (i.e. fold change) as well as p-values and false discovery rate (FDR), all calculated from four replicates by the EdgeR algorithm as discussed in the methods. Table also includes standard gene names (or acronyms) for clarity. Significantly down-regulated transcripts are in red-shaded cells whereas up-regulated sn/snoRNAs are in green-shaded cells.

Pearson correlation coefficients were calculated comparing the average $\log_2$ fold-change values in *rrp6Δ* versus WT obtained in our RNA-Seq dataset (n=4) and in a recent tiling array study by Castelnuovo *et al.* 2014 [217, 218]. The correlation coefficient when comparing these two datasets is 0.751 (Figure 15A), a value indicating that there is a strong positive correlation between the two datasets similar to the extent of correlation previously found when comparing RNA-Seq to the tiling array platform [219]. Interestingly, the highly abundant sn/snoRNA transcripts had poor correlation across the two platforms, represented as red dots in Figure 15A (values also given in Table 4). It has previously been shown that RNA-Seq has a capacity for much greater dynamic range than microarrays, which can be saturated by very abundant transcripts and may not be able to quantitatively detect transcripts expressed at low levels [56]. For this reason, we propose that the differences seen between our dataset and previously published tiling array data are due at least in part to the increased dynamic range obtained with deep transcriptome coverage obtained by our RNA-Seq experiments. Considering that sn/snoRNAs are highly abundant and that they are one of the primary cellular targets of Rrp6 activity [14-16], the ability to accurately measure the abundance of snRNAs in this instance is a distinct advantage of using RNA-Seq.

To compare the detection and quantification of extended 3' ends of a selection of snRNAs in WT versus *rrp6Δ* cells, area-under-the-curve values were calculated for snR33 and snR37 using our RNA-Seq data compared to data from the previously published tiling array dataset (Figure 15B-D, [218]). The degree of

extended transcript accumulation in *rrp6Δ* cells involved the measured ET area which was normalized against the area under the curve for the entirety of each gene transcript resulting in a percent of extended transcript value (Figure 13B). For transcripts snR37 and snR33, which have extended 3' ends, there are significantly more extended reads in the *RRP6* mutant than in wild-type in both the sequencing and array data sets (Figure 15). This agrees with the known role of Rrp6 to trim 3' ends of snRNAs after termination. However, the relative ratios of these extended transcripts to the sense transcript appear to be much higher in the tiling array data than in the sequencing data in RNA isolated from *RRP6* deletion strains. For instance, approximately 11% of snR37 transcripts appeared extended in the tiling arrays dataset as opposed to 0.3% when analyzed by RNA-Seq (Figure 15B). RNA sequencing read count values in our study span 5 orders of magnitude while the tiling array data only covers 3-4 orders of magnitude, likely causing a loss of accurate detection of highly abundant fully processed sn/snoRNAs (compare Figure 15C, D). The highly abundant processed snR37 or snR33 peaks are clearly distinguished from extended products using RNA-Seq (Figure 15). Our data suggests that the relative effects of the loss of *RRP6* on the steady state levels of these extended transcripts is much less dramatic than previously described and that quantitation of steady state levels of the processed snRNA upon loss of *RRP6* was problematic due to limited dynamic range of the microarray approach.

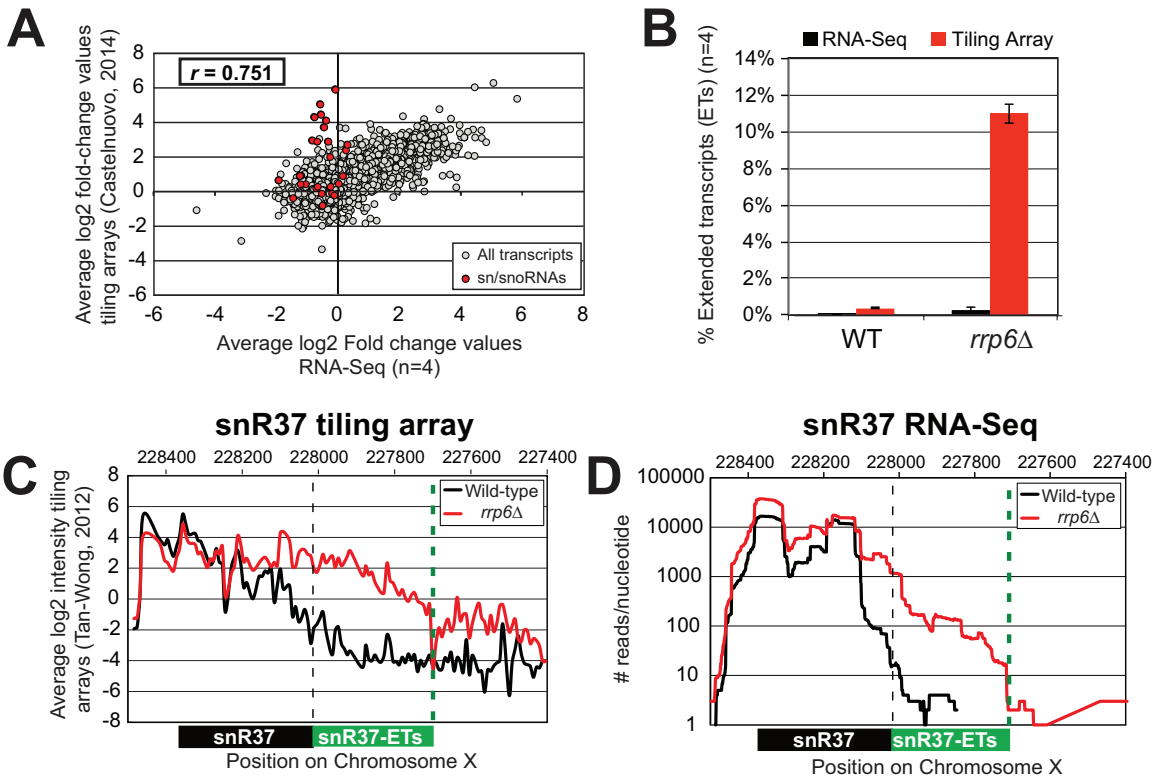

Figure 15: Comparison of highly abundant sn/snoRNA transcripts from *rrp6Δ* and WT strains obtained through tiling array or RNA sequencing. (A) Comparison of a recent *rrp6Δ* tiling array dataset [51] and RNA-Seq data collected in this study by plotting the average $\log_2$ ratio values (*rrp6Δ* / WT) for all transcripts. All annotated transcripts included in both data sets are represented by gray circles. All sn/snoRNAs are highlighted in red circles. The Pearson correlation coefficient (*r*) between these datasets = 0.751, indicating a modest positive correlation between the two data sets. (B) Area under the curve calculation for snR37 extended transcripts (labeled as "ET") from WT and *rrp6Δ* strains. The area calculated for snR37-ET annotation was divided by the area calculated for the entire snR37 to snR37-ET annotation to calculate the percentage of the entire transcript area located in the snR37-ET region (see diagram under (C) and (D) for locations of annotations). Values are shown as averages ± standard deviations for sequencing data are in black, tiling array data is in red. (C) Previously published tiling array intensity values [59] at snR37 using probe mid-position (8 nt apart), for comparison with (D) mapped reads from RNA-Seq data at snR37 (single nt resolution). Tiling array data represented in $\log_2$ scale. RNA-Seq data represented in $\log_{10}$ scale. In both graphs, WT expression is in black, and *rrp6Δ* is in red. Locations of annotated transcripts and the direction of transcription are indicated below the charts.

Transcript class	Standard gene name	RRP6 _KO_ vs WT log2_FC	RRP6 _KO vs WT p-value	RRP6 _KO vs_WT FDR
sn/snoRNA	SNR81	-2.860386521	4.85E-14	4.14E-12
sn/snoRNA	SNR44	-2.230899649	2.77E-07	3.50E-06
sn/snoRNA	SNR35	-1.928146267	2.05E-10	6.41E-09
sn/snoRNA	SNR56	-1.523807602	6.39E-05	0.000413774
sn/snoRNA	SNR86	-1.462375099	8.21E-07	9.17E-06
sn/snoRNA	SNR65	-1.430941739	9.68E-06	7.85E-05
sn/snoRNA	SNR78	-1.365952969	5.82E-04	0.002888894
sn/snoRNA	SNR64	-1.364898817	2.01E-05	0.000148358
sn/snoRNA	SNR84	-1.236072278	0.000125125	0.000748372
sn/snoRNA	SNR33	-1.215933047	9.83E-05	0.000603253
sn/snoRNA	SNR66	-1.195949546	9.14E-05	0.000563868
sn/snoRNA	SNR59	-1.107521289	0.014237604	0.046625789
sn/snoRNA	SNR189	-1.048743348	0.001930926	0.008499109
sn/snoRNA	SNR191	-0.896795436	0.020416846	0.062502573
sn/snoRNA	snR61	-0.849755476	0.013504106	0.044647673
sn/snoRNA	SNR34	-0.833187075	0.014336742	0.046878903
sn/snoRNA	SNR71	-0.765950611	0.015401095	0.04976118
sn/snoRNA	SNR87	-0.675373826	0.016687617	0.053235664
sn/snoRNA	SNR46	-0.661274027	0.02844551	0.082096313
sn/snoRNA	SNR50	-0.628210692	0.029626572	0.084924341
sn/snoRNA	SNR31	-0.713728032	0.037316016	0.102087548
sn/snoRNA	SNR14	0.711639383	0.035216361	0.09758774
sn/snoRNA	SNR6	0.82703613	0.022085331	0.066560823
sn/snoRNA	SNR57	0.881604448	0.006580374	0.024512926
sn/snoRNA	SNR75	1.069032702	0.004722081	0.01845638
sn/snoRNA	SNR42-A	1.878888595	2.73E-08	4.71E-07
sn/snoRNA	SNR87-A	2.31736285	1.76E-09	4.31E-08
sn/snoRNA	SNR39B	-0.909027838	0.039039925	0.106084238
sn/snoRNA	SNR40	-0.591588028	0.052996397	0.133670752
sn/snoRNA	SNR36	-0.635111023	0.058982769	0.145245968
sn/snoRNA	SNR80	-0.693626771	0.0740974	0.17212582
sn/snoRNA	SNR128	-0.574195832	0.118859877	0.243317703
sn/snoRNA	SNR42	-0.549369317	0.06806735	0.161660948
sn/snoRNA	SNR52	-0.537526624	0.120245375	0.245605501
sn/snoRNA	SNR8	-0.532345301	0.142159353	0.278487816
sn/snoRNA	SNR79	-0.527744145	0.149737192	0.287773936
sn/snoRNA	SNR18	-0.519381528	0.085166497	0.190933516
sn/snoRNA	SNR47	-0.517254433	0.143937233	0.280475953
sn/snoRNA	SNR83	-0.505712913	0.095794227	0.208546845
sn/snoRNA	SNR60	-0.487155005	0.218627515	0.374429324
sn/snoRNA	SNR9	-0.452971932	0.198883562	0.349702151

Transcript class	Standard gene name	RRP6 _KO_ vs WT log2_FC	RRP6 _KO vs WT p-value	RRP6 _KO vs_WT FDR
sn/snoRNA	SNR77	-0.444248953	0.149459373	0.287404332
sn/snoRNA	SNR161	-0.423553015	0.179671466	0.326981965
sn/snoRNA	SNR43	-0.416456202	0.226726278	0.383525692
sn/snoRNA	SNR32	-0.383845798	0.2510801	0.412302527
sn/snoRNA	SNR55	-0.374299562	0.296846966	0.461555507
sn/snoRNA	SNR38	-0.355406127	0.340398658	0.504939698
sn/snoRNA	SNR190	-0.324843091	0.361747053	0.525783972
sn/snoRNA	SNR49	-0.318735612	0.313881222	0.479209424
sn/snoRNA	SNR54	-0.29666824	0.338839999	0.50344225
sn/snoRNA	SNR3	-0.269967022	0.452510653	0.603581053
sn/snoRNA	SNR70	-0.26471889	0.394224359	0.555372791
sn/snoRNA	SNR63	-0.261437558	0.403192231	0.562039045
sn/snoRNA	SNR53	-0.258692147	0.477901804	0.624237615
sn/snoRNA	SNR30	-0.24430417	0.434322121	0.588828215
sn/snoRNA	SNR76	-0.229565876	0.418323974	0.574389084
sn/snoRNA	SNR24	-0.219193244	0.503883819	0.645965476
sn/snoRNA	SNR37	-0.198166563	0.533192777	0.669193761
sn/snoRNA	SNR4	-0.125202438	0.739787893	0.827326293
sn/snoRNA	SNR41	-0.122609039	0.665292218	0.771659389
sn/snoRNA	SNR48	-0.090758158	0.808692987	0.875736692
sn/snoRNA	SNR45	-0.083113552	0.795439103	0.867438943
sn/snoRNA	SNR69	-0.079138447	0.782649698	0.859253082
sn/snoRNA	SNR67	-0.056167085	0.856106882	0.907683924
sn/snoRNA	SNR72	-0.055840218	0.859360009	0.909698334
sn/snoRNA	SNR11	-0.051426797	0.89548336	0.933960656
sn/snoRNA	SNR7	-0.038192099	0.903347021	0.93988355
sn/snoRNA	SNR82	0.020484443	0.948940617	0.970148067
sn/snoRNA	SNR17A	0.055254753	0.862752868	0.911471967
sn/snoRNA	SNR17B	0.072352447	0.81444406	0.880168674
sn/snoRNA	SNR58	0.092225721	0.760947806	0.843045229
sn/snoRNA	SNR39	0.106069082	0.71306577	0.806839679
sn/snoRNA	SNR62	0.15611664	0.674808634	0.778214861
sn/snoRNA	SNR19	0.158597522	0.586520686	0.711472945
sn/snoRNA	SNR51	0.206204667	0.507219697	0.64875741
sn/snoRNA	SNR68	0.243528938	0.498762767	0.641357274
sn/snoRNA	SNR5	0.259200468	0.481388595	0.627416647
sn/snoRNA	SNR13	0.279318313	0.418252867	0.574389084
sn/snoRNA	SNR85	0.306657293	4.10E-01	5.68E-01
sn/snoRNA	SNR74	0.374300957	2.84E-01	4.47E-01

Table 8: Differential expression information for the sn/snoRNAs. Table includes differential expression data expressed in $\log_2$ *rrp6Δ*/WT ratio (i.e. fold change), as well as p-values and false discovery rate (FDR), all calculated from four replicates by the EdgeR algorithm as discussed in the methods. Significantly down-regulated sn/snoRNAs are in red-shaded cells whereas up-regulated sn/snoRNAs are in green-shaded cells.

2.2. Analysis of snRNA termination in *RRP6* deletion mutants

Loss of *RRP6* results in improper 3'-end processing of multiple sn/snoRNAs. Such 3'-end processing defects could cause instability of snRNAs that leads to changes in their overall expression levels in cells [57,58]. Based on our differential expression analysis, 24 of the 78 sn/snoRNAs show decreased expression of at least 1.5-fold ($\log_2 \geq -0.6$) in *rrp6Δ* cells versus WT cells with false discovery rates of ≤ 0.1 (Table 8). To better understand the role of Rrp6 in global snRNA 3' end processing, the length and intensity of snRNA extended transcripts was analyzed in *rrp6Δ* cells through manual annotation of the extended transcripts (annotated as "ETs") observed in our RNA-Seq dataset. The length of snRNA transcript extension was compared to the annotated length of Nrd1-unterminated transcripts (NUTs) and cryptic unstable transcripts (CUTs) [73, 214]. Additionally, we performed ChIP-exo according to established protocols to generate high-resolution maps of RNAPII localization throughout the yeast genomes in WT and *RRP6* deletion cells with Rpb3-FLAG strains made for this study (Figure 15B) [211]. It was recently reported that NUT annotations were significantly longer than CUT annotations, suggesting that Rrp6 is required for 3'-end processing of NNS terminated transcripts but is not directly involved in NNS termination [214]. Comparison of our datasets from RNA sequencing and ChIP-exo to NUT annotations as well as direct comparison of specific transcripts in *rrp6Δ* cells to Nrd1 mutant cells via northern analysis allowed us to discover a regulatory role for Rrp6 on NNS-dependent RNAPII termination.

Kim *et al* previously reported [20] that some snRNAs are significantly longer in Nrd1-depleted cells than in *rrp6Δ* cells, and our dataset confirms these findings. It was found that there is a large extension for snR13 in *SEN1* and *NRD1* mutant strains, with the snR13 transcript extended across the transcribed region of *TRS31* [20]. This extension at snR13 has also been reported in *SSU72* mutant strains [193, 218, 220]. NUT0167 was annotated as an Nrd1-dependent unterminated transcript at snR13 following Nrd1 nuclear depletion. Kim *et al.* also identified both normal length and an extended snR13 transcript in *rrp6*Δ cells, which corresponds to the pre-snRNA transcript that is not correctly processed in the absence of Rrp6 by northern blotting. For direct comparison of our dataset to previous work, we validated our RNA-Seq and ChIP-exo results by northern blot analysis using a strand specific probe against the processed version of snR13 (Figure 16). The extended transcript detected in Nrd1 nuclear depletion experiments was annotated as NUT0167 and is 1378 nt longer than the pre-snR13 transcript observed in our *rrp6*Δ RNA-Seq samples (Figure 16A). In agreement, we observed a transcript approximately 1500 nt in length by northern blotting in the *NRD1*-temperature sensitive *(ts)* mutant (*nrd1Δ151–214* that lacks the Nab3 interaction domain [19, 221, 222]) following 30 or 60 minute heat shock to inactivate the *NRD1* mutant. Long extended transcripts were also observed in *ssu72 TOV*, a "terminator override" mutant previously described as deficient in NNS-dependent termination (kindly provided from the Reines lab) [223]. However, no such read-through transcripts were observed in the *rrp6*Δ cells by RNA-Seq or by northern blotting, only the unprocessed pre-snR13 (Figure 16A,

D). Supporting the RNA transcript data, we see no shift in RNAPII localization in this region as detected by Rpb3-FLAG ChIP-exo sequencing (Figure 16B). Previous studies to identify Nrd1 binding sites in target RNAs using PAR-CLIP sequencing revealed a strong Nrd1 signal just downstream of the annotation for the mature snR13 transcript, supporting the role of the NNS pathway in termination in this region ([50], Figure 16B). No significant changes were observed in the expression levels of snR13 or *TRS31,* which also suggests that snR13 RNA is correctly terminated in *rrp6*Δ cells (Figure 15C). Taken together, these data clearly show that Rrp6 is not required for termination of snR13 transcription by the NNS pathway, which is in agreement with previous findings [20].

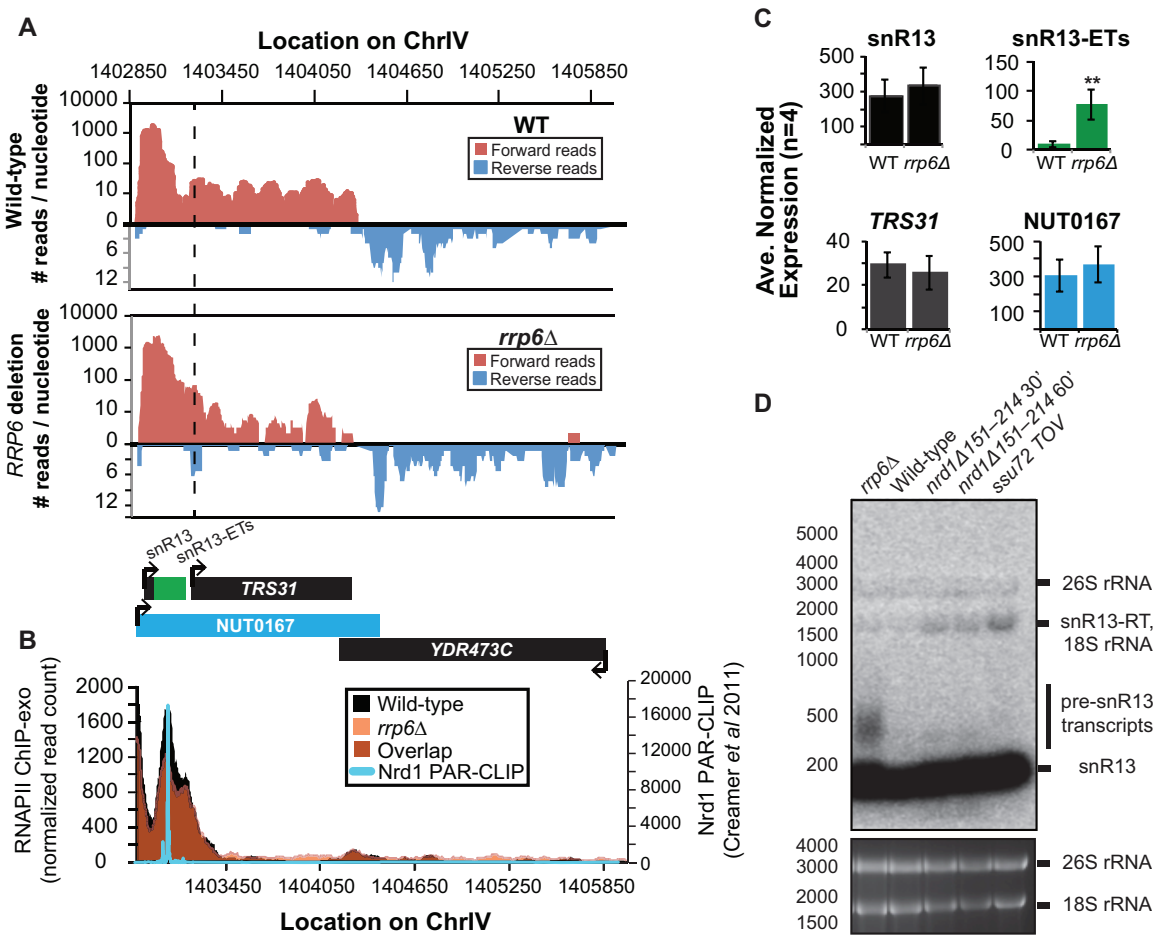

Figure 16: Transcription termination of the C/D box small nucleolar RNA snR13 does not require Rrp6. (A) Graphical representation of strand-specific RNA-Seq reads mapped to snR13-*YDR473C* region. Reads mapped to the positive strand are on top in red, whereas reads mapped to the negative strand are on the bottom in blue. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Processed length of snRNAs and mRNAs are in black, snRNA-extended transcripts, including pre-snRNAs and termination read-through products, are in green (labeled "ETs"), NUTs are in aqua, and arrows indicate annotated transcription start site. The dotted black line marks the transcription start site (TSS) of *TRS31*. (B) Rpb3-FLAG localization as determined by ChIP-exo sequencing reads mapped to the same region and aligned to (A). WT reads are in black, and *rrp6Δ* are in orange. Nrd1 binding sites as determined by PAR-CLIP from [50] are shown for comparison in aqua. (C) Average normalized read counts ± standard deviations for significantly altered transcripts in *rrp6Δ* cells versus WT cells (n=4). Two stars indicate a p-value of <0.01 as determined by an unpaired, two-tailed student's t-test. The colors of the bars in each graph correspond to the color representing the related annotation. (D) Strand-specific northern blot analysis using a 5' end labeled DNA oligonucleotide probe specific to the processed region of snR13

directly comparing *rrp6Δ* to mutants known to be defective in NNS-dependent termination. The temperature sensitive *nrd1Δ151-214* strain was grown at 30C overnight, diluted to an OD_{600} of 0.5, and grown at 37C for 30 minutes or 60 minutes. The *ssu72 TOV* strain has been previously shown by [223] to bypass NNS-dependent termination at the *IMD2* locus.

In contrast to previous conclusions that Rrp6 is dispensable for snRNA termination, our data indicates that a subset of snRNA transcripts require Rrp6 for proper RNAPII termination *in vivo* [20, 214]. As shown in Figure 17A, the apparent length of snR3 in *rrp6Δ* cells as revealed by RNA-Seq is 1396 nt whereas the annotation for NUT0426 is 3363 nt long (Figure 17A). However, the Rrp6-dependent snR3 extension is 246 nt longer than the three tandem CUTs (CUT221, 222, and 223) previously annotated from tiling array data [218]. All three CUTs were significantly up-regulated in *rrp6Δ* cells as expected (Figures 17B, 18).

Of interest, a comparison of the 4tU-Seq data used to annotated the NUTs with our RNA-Seq data clearly shows a difference in the length of the Nrd1-unterminated transcript NUT0426 resulting from snR3 read-through than that observed upon deletion of *RRP6* (Figure 18). It is interesting to note that the overall read count for the snR3 read-through product in the 4tU-Seq dataset decreases dramatically just 3'- to the read-through observed in *rrp6Δ* cells, suggesting that the majority of transcripts are terminated at the location indicated in the *rrp6Δ* RNA-Seq (Figures 17, 18). ChIP-exo analysis of RNAPII localization in this region revealed an increase in RNAPII density just downstream of snR3 in the *rrp6Δ* cells when compared to WT (Figure 17B, arrow 1). This increase in downstream RNAPII corresponds to the increased level of RNA detected downstream of snR3 in Figure 17A past the pre-snRNA transcript and previously mapped Nrd1 and Nab3 binding sites located ~400bp from the start of the snR3 transcript [50, 60]. Although the RNAPII density decreases just before the 3'-end

of the snR3-ETs annotation, it remains slightly elevated in *rrp6Δ* cells compared to WT throughout the NUT0426 annotation region, indicated by arrow #2 (Figure 17B).

Differential expression analysis also revealed highly significant down-regulation of the snR3 convergent, "tail-to-tail," gene *YJR129C* likely as a result of extended snR3 transcripts in *rrp6Δ* cells (Figure 16C). These data suggest that extension of snR3 causes transcription interference at *YJR129C* as was also reported in Nrd1 nuclear depletion experiments [46]. To definitively determine the length of the snR3 transcripts and compare *rrp6*Δ cells directly to an *NRD1* mutant, we also performed northern blot analysis with a probe that recognizes the short, processed snR3. A transcript that is approximately 4000 nt long was detected in *rrp6Δ, nrd1-ts*, and the *ssu72 TOV* mutants (Figure 17D). This transcript is approximately the length expected for NUT0426. We also detected two shorter transcripts approximately 500 and 1000 nt long in the *nrd1-ts* and *rrp6Δ* cells. Together, these data clearly show that Rrp6 is required for proper RNAPII termination at snR3.

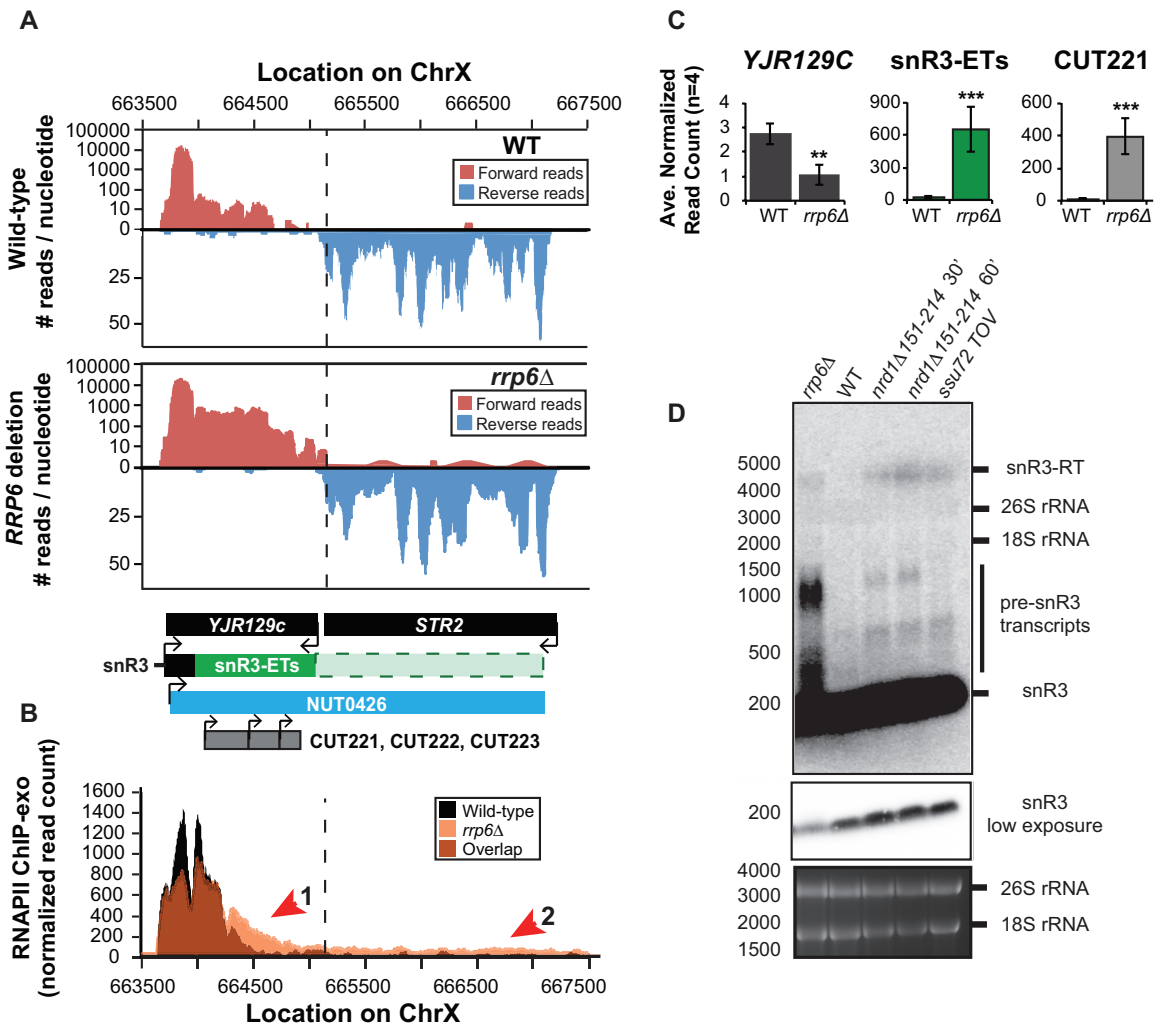

Figure 17: The H/ACA box small nucleolar RNA snR3 requires Rrp6 for efficient termination. (A) Graphical representation of strand-specific RNA-Seq reads mapped to snR3-*STR2* region. Reads mapped to the positive strand are on top in red, while reads mapped to the negative strand are on the bottom in blue. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Processed length of snRNAs and mRNAs are in black, snRNA-extended transcripts, including pre-snRNAs and termination read-through products, are in green (labeled "ETs"), NUTs are in aqua, CUTs are in gray, and arrows indicate annotated transcription start site. The dotted black line marks the transcription start site (TSS) of *YJR129C*. (B) Rpb3-FLAG localization as determined by ChIP-exo sequencing reads mapped to the same region and aligned to (A). WT reads are in black, and *rrp6Δ* are in orange. (C) Average normalized read counts ± standard deviations for significantly altered transcripts in *rrp6Δ* cells versus WT cells (n=4). Two stars indicate a p-value of <0.01, and three stars indicate a p-value of <0.001 as determined by an unpaired, two-tailed student's t-test. The colors of the bars in each graph correspond to the color representing the related annotation. (D)

Strand-specific northern blot analysis using a 5’ end labeled DNA oligonucleotide probe specific to the processed region of snR3 directly comparing *rrp6Δ* to mutants known to be defective in NNS-dependent termination.

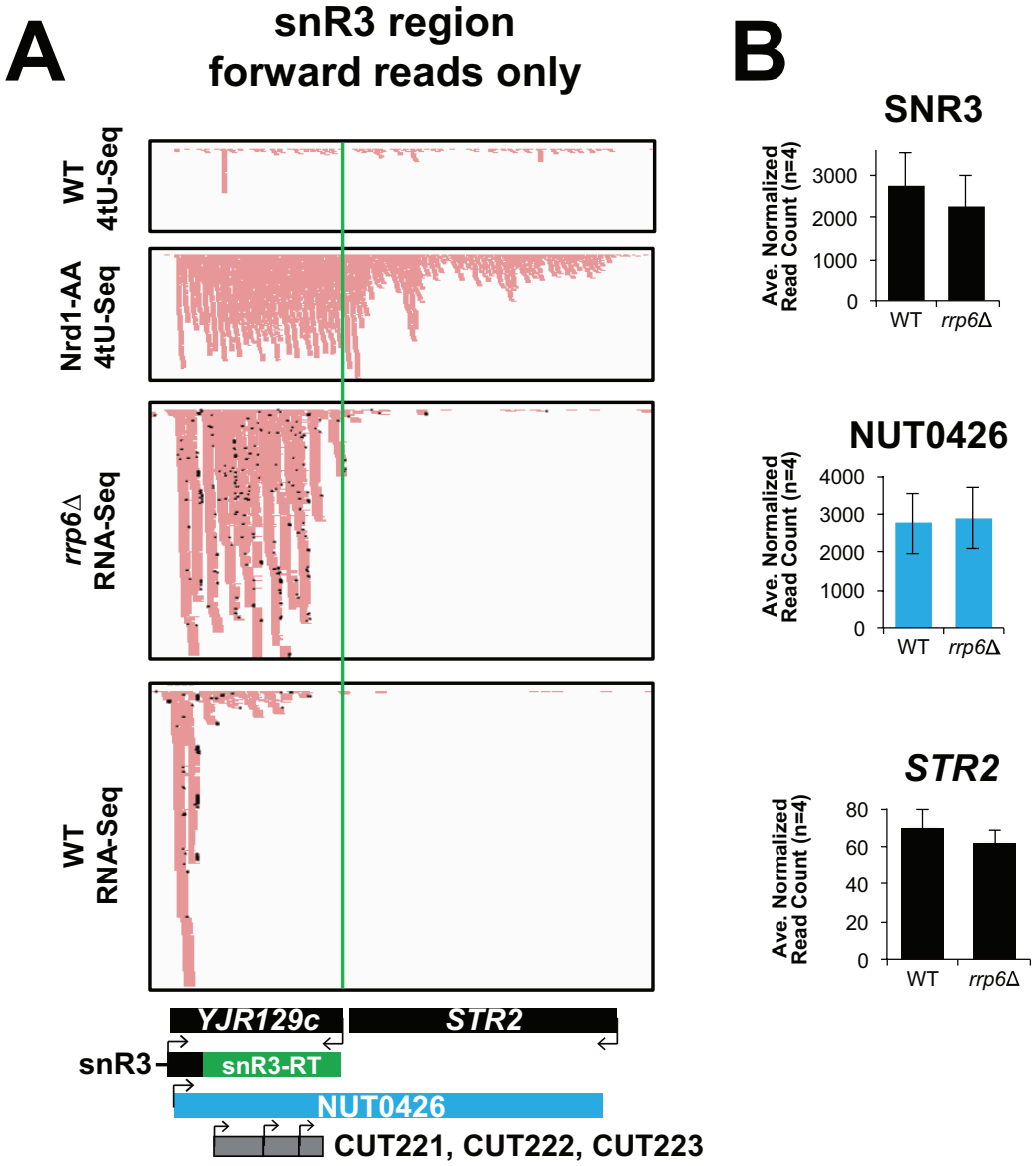

Figure 18: Comparison of *rrp6Δ* RNA-Seq reads at snR3 to published 4tU-Seq data for NUTs. (A) Mapped reads obtained from our in-house alignment (see methods section) of 4tU-Seq data from [214] (top two panels) and our RNA-Seq data (bottom two panels) at snR3-*STR2* region, forward strand reads only. The location and transcription direction of all annotations within this region are diagrammed below. Processed length of snRNAs and mRNAs are in black, snRNA-extended regions are in green (labeled "ET"), NUTs are in blue, CUTs are in gray, and arrows indicate annotated transcript state site and direction of transcription. The dotted green line marks the 3' end of the extended snR3 annotation in *rrp6Δ* cells. (B) Average normalized read counts ± standard deviations for transcripts in this region that are not significantly changed in *rrp6Δ* cells versus WT (n=4). The colors of the bars correspond to the color representing the annotation.

2.3. *RRP6* deletion results in extension of snRNA transcripts leading to down-regulation of neighboring genes

Several genes downstream of snRNAs transcribed from the opposite DNA strand were found to be significantly down-regulated when transcript extension was high in *rrp6Δ* cells. We focused on snRNAs with “tail-to-tail” or convergent genes because the strand specific sequencing data allows us to distinguish between reads mapped to the two different transcripts. As mentioned in the example of snR3, the expression of neighboring downstream gene *YJR129C* is down almost three-fold (Figure 17C). The detectable snR3 transcript maps beyond the transcription start site (TSS) for *YJR129C* suggesting that transcription interference is the likely mode of *YJR129C* repression as diagrammed in Figure 19A. Additional examples of snRNA termination defects causing transcription interference at downstream genes are shown in Figures 15 and 16. As shown in Figure 15, snR11 has a downstream gene transcribed on the opposite strand, *CMC4* (Figure 19C). snR11 has a long region of RNA-Seq reads in *rrp6Δ* cells that extends beyond the *CMC4* TSS, and *CMC4* expression is decreased by more than half (Figure 19C). However, northern blot analysis with a single-stranded oligonucleotide probe detecting the short, processed snR11 does not show read-through transcripts >1000 bp in any of the samples, including the mutants known to have defective Nrd1 termination (Figure 19B). Since there is no detectable read-through snR11 transcript in the *nrd1-ts* mutant, the transcript encoded in the region annotated “snR11-ETs” and “NUT0607” is likely initiated at start site downstream of the snR11 sequence recognized by our

probe. The lack of highly extended transcripts, readily distinguishable by northern blot, shows the limitation of using short sequencing reads for transcript annotation (Figure 19). Interestingly, Rpb3-FLAG ChIP-exo data shows that the peak of RNAPII localization at snR11 is extended farther 3' in *rrp6Δ* cells that in WT cells indicated by arrow #1, and this extension does indeed overlap the TSS of *CMC4* (Figure 19D). RNAPII occupancy quickly decreases just beyond the *CMC4* TSS but remains higher than in WT cells. The increase in RNAPII localization in *rrp6Δ* cells downstream of the native termination site for snR11 suggests that loss of *rrp6Δ* decreases the efficiency of snR11 termination leading to a small degree of read-through that is sufficient to cause transcription interference at *CMC4*.

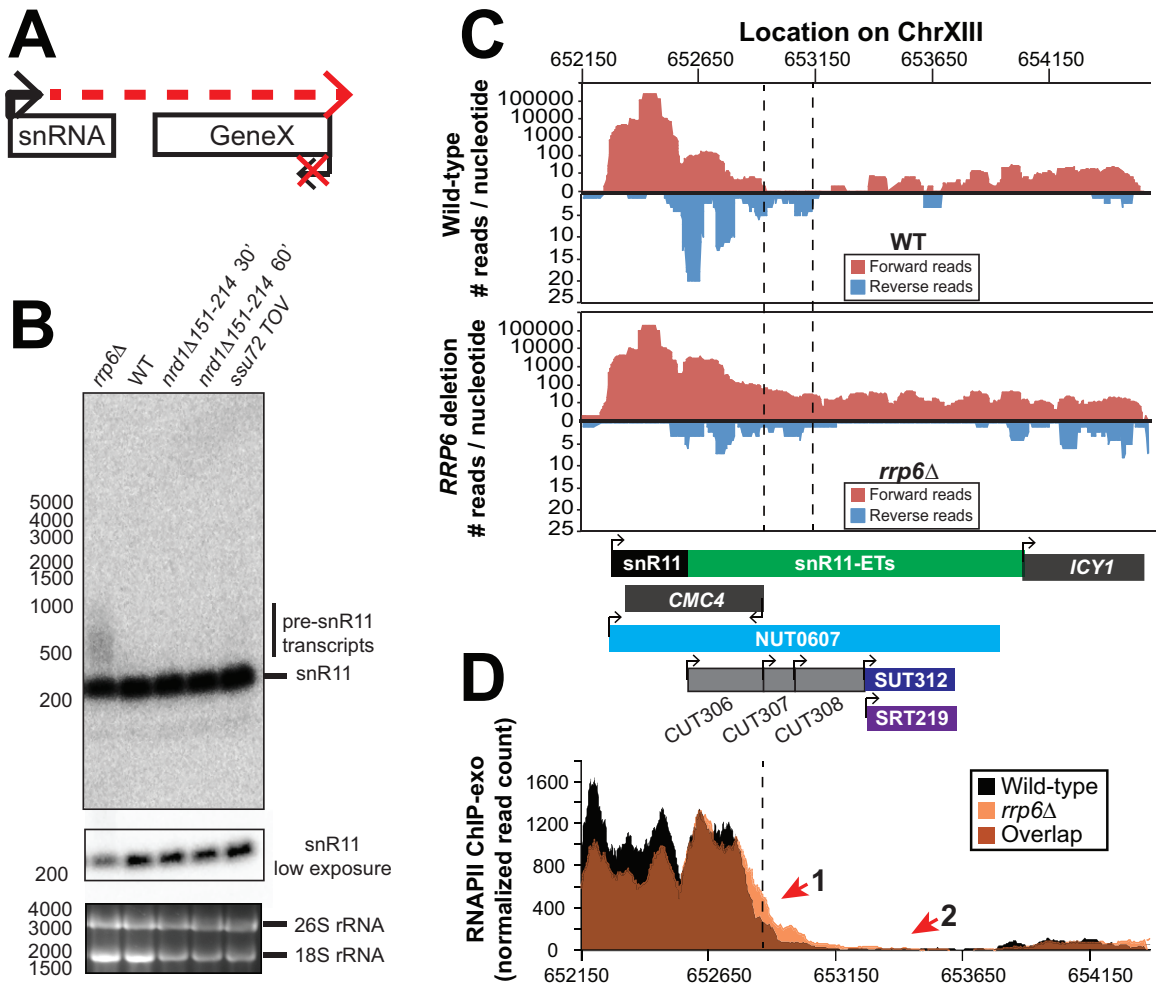

Figure 19: Transcription termination of the H/ACA box small nucleolar RNA snR11 shifts toward the 3'-end in *rrp6Δ* cells. (A) Diagram showing proposed mechanism where down-regulation of GeneX results from faulting termination of snRNA in *rrp6Δ* cells. Decreased NNS-termination at select snRNAs results in longer transcribed region, extending over transcription start site of downstream convergent gene, GeneX. The hypothesized resulting increased localization of the transcription machinery interferes with initiation at the TSS of GeneX (indicated with a red 'X'). (B) Strand-specific northern blot analysis using a 5'-end labeled DNA oligo probe specific to the processed region of snR11 directly comparing *rrp6Δ* to mutants known to be defective in NNS-dependent termination. (C) Graphical representation of strand-specific RNA-Seq reads mapped to snR11-*ICY1* region. Reads mapped to the positive strand are on top in red, while reads mapped to the negative strand are on the bottom in blue. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Processed length of snRNAs and mRNAs are in black, snRNA-extended transcripts, including pre-snRNAs and termination read-through products, are in green (labeled "ETs"), NUTs are in aqua, CUTs are in gray, SUTs are in dark blue, SRTs are in purple, and bent

arrows indicate direction of the TSS. The dotted black lines mark the transcription start sites (TSS) of *CMC4*. (D) Rpb3-FLAG localization as determined by ChIP-exo sequencing reads mapped to the same region and aligned to (C). WT reads are in black, and *rrp6Δ* are in orange.

A substantial read-through transcript (1442 nt) was detected at snR71, a C/D box small nucleolar RNA gene, in *rrp6Δ* cells that extends well beyond the TSS of downstream gene *LIN1,* which is significantly down-regulated (Figure 20A, C) [224, 225]. There is a 2281 nt long NUT annotated at this locus (NUT0349) that overlaps with a second downstream gene, *REC104*, which is on the same strand as snR71 (Figure 20A). With strand-specific sequencing, it cannot be absolutely determined if reads mapped to the *REC104* locus are from *REC104* transcripts, from a much longer snR71 read-through transcript, or from a combination of both. To answer this question and compare the effects of *RRP6* deletion to mutants with defective NNS-dependent termination directly, we performed northern blot analysis with a short probe specific to the processed snR71 transcript (Figure 20D). We see a striking band at approximately 1700 nt in *rrp6Δ, nrd1-ts, and Ssu72 TOV* cells. This suggests that a percentage of RNAPIIs transcribing snR71 in all of these mutant strands can escape the NNS-dependent termination pathway and are hence terminated much farther downstream (Figure 20). This is supported by the Rpb3-FLAG ChIP-exo data that shows an increase in RNAPII localization downstream of snR71 in *rrp6Δ* cells compared to WT cells along the full length of the NUT0349 annotation (Figure 20B). Interestingly, the expression level of the annotated mature snR71 region is also significantly decreased (Figure 20C), suggesting there may be some instability or other defects caused by improper 3'-processing and/or termination caused by loss of Rrp6.

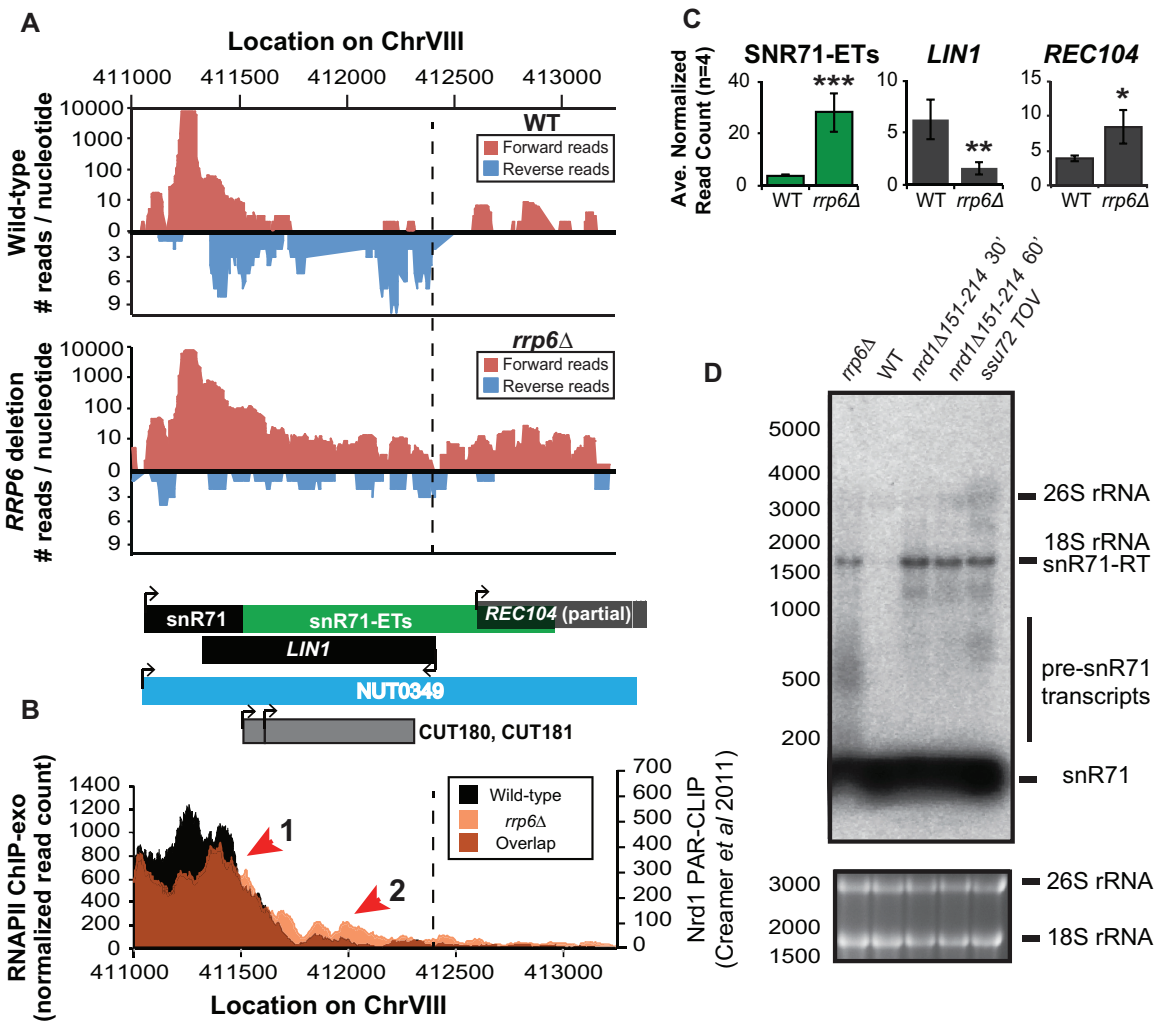

Figure 20: Rrp6 is required for NNS-dependent termination and RNA processing of the snR71 transcript. (A) Graphical representation of strand-specific RNA-Seq reads mapped to snR71-extended region. Reads mapped to the positive strand are on top in red, while reads mapped to the negative strand are on the bottom in blue. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Processed length of snRNAs and mRNAs are in black, snRNA-extended transcripts, including pre-snRNAs and termination read-through products, are in green (labeled "ETs"), NUTs are in aqua, CUTs are in gray, and bent arrows indicate direction of the TSS. The dotted black line marks the transcription start site (TSS) of *LIN1*. (B) Rpb3-FLAG localization as determined by ChIP-exo sequencing reads mapped to the same region and aligned to (A). WT reads are in black, and *rrp6Δ* are in orange. (C) Average normalized read counts ± standard deviations for significantly altered transcripts in *rrp6Δ* cells versus WT cells (n=4). One star indicates a p-value of <0.05, two stars indicate a p-value of <0.01, and three stars indicate a p-value of <0.001 as determined by an unpaired, two-tailed student's t-test. The colors of the bars in each graph correspond to the color representing the related

annotation. (D) Strand-specific northern blot analysis using a 5’ end labeled probe specific to the processed region of snR71 directly comparing *rrp6Δ* to NNS termination mutants.

Overall these data show that multiple transcripts downstream of snRNAs have significant changes in expression following deletion of *RRP6.* ChIP-exo analysis of RNAPII (Rpb3-FLAG) clearly show that RNAPII terminates downstream of its normal stopping point in WT cells at specific snRNAs leading to mislocalization and causing transcription interference at downstream genes (Figures 17, 19, 20). In this model, ineffective termination of the snRNAs results in transcription interference at the downstream gene as previously described following inactivation of Nrd1 (Figure 19A [46]). These findings, taken together with the similar lengths of Rrp6-dependent snRNA transcripts and Nrd1-dependent transcript annotations (NUTs), support the hypothesis that Rrp6 serves as important regulator of NNS-dependent termination at specific sn/snoRNAs.

2.4. *RRP6* is required for proper RNAPII termination of NNS-dependent regulatory non-coding RNAs

In addition to extension of snRNA transcripts, changes in expression and/or apparent length of other previously described noncoding RNAs, such as CUTs, SUTs, and NUTs, were also detected with corresponding down-regulation of neighboring genes in cells lacking Rrp6 suggesting transcription interference. A number of these transcripts have been well-characterized as being extended following depletion or genetic inactivation of NNS components but not the exonuclease *RRP6*.

An early and well-characterized example of a gene that is regulated by NNS-dependent termination is the *NRD1* gene [43, 145, 171]. There are a cluster of multiple Nrd1 and Nab3 binding sites in the 5'-UTR of the *NRD1* mRNA leading to early termination of the transcript and autoregulation of *NRD1* expression levels through a mechanism that also requires Sen1 [43, 49, 50, 171]. In expression analysis of the RNA-Seq data, we see nearly a 2-fold increase in the expression of the *NRD1* transcript in *rrp6Δ* cells, as well as an increase in CUT320, a noncoding transcript near the *NRD1* promoter on the opposite strand (Figure 21A). Since Rrp6 is responsible for the degradation of the short *NRD1* transcript, this increase in expression could be attributed to stabilization of terminated transcript. Alternatively, increased expression of *NRD1* in *rrp6Δ* cells could suggest that the early termination sites are not being utilized by the NNS pathway as often in the *rrp6Δ* cells, leading to transcription of the full-length mRNA. Analysis of RNAPII occupancy at the *NRD1* gene in the ChIP-exo dataset can distinguish these two possibilities (Figure 21B). In WT cells, there are several peaks of RNAPII at the 5'-end of the *NRD1* gene that quickly decreases to low levels along the *NRD1* transcribed region in agreement with previous findings in an *NRD1* mutant [171]. In fact, comparison of the ChIP-exo data with Nrd1 RNA binding sites mapped by PAR-CLIP reveal that the majority of RNAPII terminates just downstream of the final Nrd1 binding site in WT cells (Figure 21B). In *rrp6Δ* cells, the intense peaks of RNAPII localization at the 5'-UTR are shifted 3' and RNAPII localization is higher along the entire length of the gene, including a higher peak at the poly-A dependent termination site of the full-

length transcript (Figure 21B). This supports the hypothesis that the NNS-termination pathway is not terminating the short 5'-UTR transcript efficiently in *rrp6Δ* cells and that RNAPII continues down the length of the gene producing the full transcript and increasing overall *NRD1* expression levels (Figure 21). The requirement for multiple Nrd1-Nab3 RNA binding sites at *NRD1* when compared to RNAPII ChIP-exo data suggest that NNS-dependent termination requires multiple Nrd1 and/or Nab3 binding sites to effectively terminate RNAPII as has been previously proposed (Figure 21B) [40, 50]. The 3' shift in RNAPII localization observed at *NRD1* supports this hypothesis.

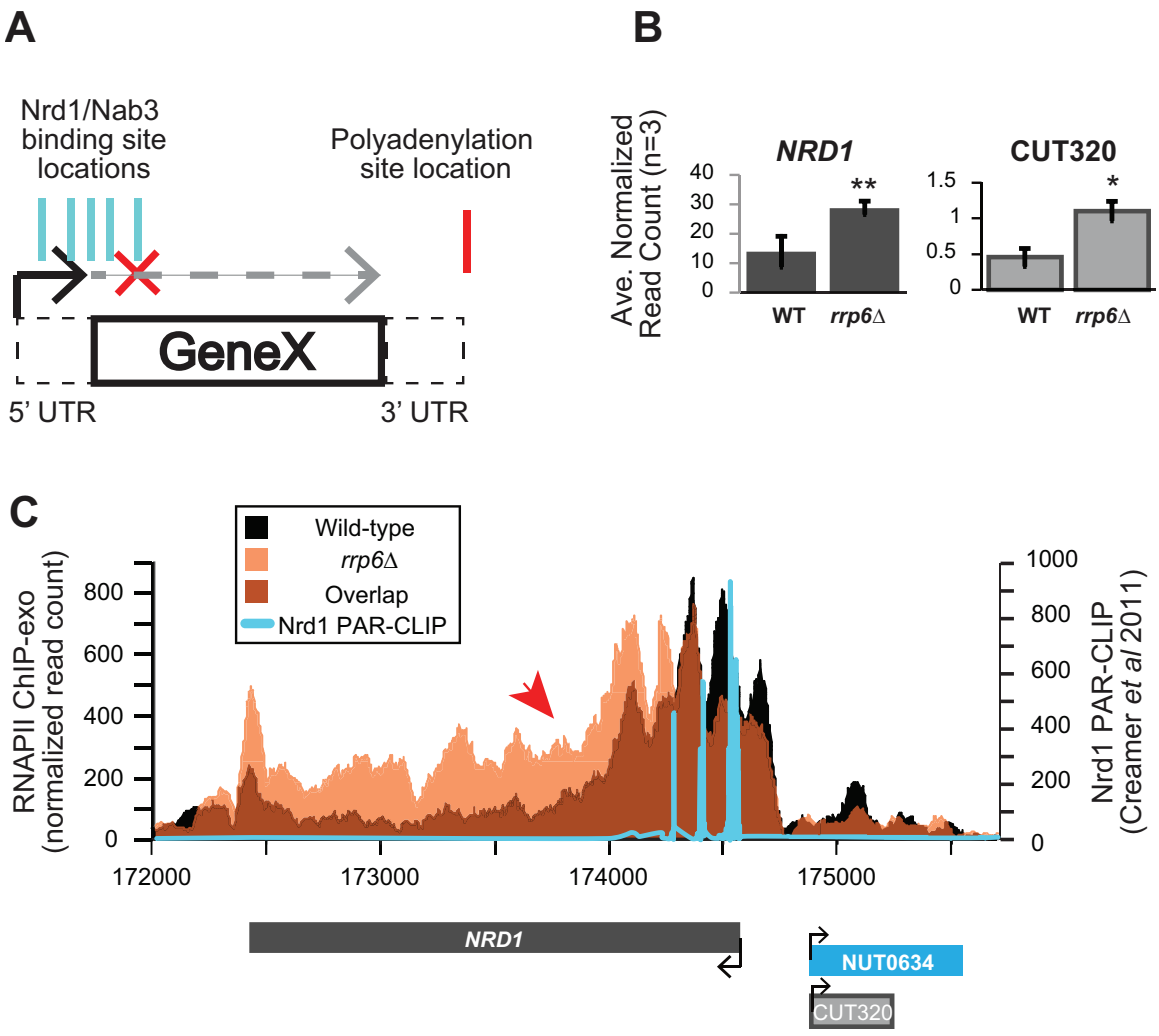

Figure 21: Early termination of *NRD1* transcription requires Rrp6 for efficient RNAPII termination. (A) Average normalized read counts ± standard deviations for *NRD1* and CUT320 transcripts in *rrp6Δ* cells versus WT (n=3). One star indicates a p-value of <0.05 and two stars indicate a p-value of <0.01. (B) Rpb3-FLAG localization as determined by ChIP-exo sequencing reads mapped to the *NRD1* region. WT reads are in black, and *rrp6Δ* are in orange. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Length of mRNAs including untranslated regions are in black, CUTs are in gray, and NUTs are in aqua. Nrd1 binding sites as determined by PAR-CLIP from [50] are shown in aqua for comparison. The scale for Nrd1 PAR-CLIP data is shown to the right.

Changes in RNAPII occupancy downstream of known NNS-dependent early termination (also known as attenuated) targets was also observed at *HRP1* and *SRG1-SER3* (Figures 22, 23). The RNAPII occupancy at *HRP1* shows similar changes as seen at *NRD1* with the majority of RNAPII terminating early in WT cells while showing a 3'- shift (arrow #1) and persistence through the gene (arrow #2) in *rrp6Δ* cells (Figure 22A). RNA-Seq analysis also revealed an increase in downstream *HRP1* RNA levels relative to WT (Figure 22B). Another well-characterized example of upstream noncoding RNA regulation occurs at the *SRG1-SER3* region. *SRG1* (*SER3* regulatory gene 1) is a known non-coding RNA whose transcription down-regulates expression of *SER3* [168]. *SRG1* RNA is bound by both Nrd1 and Nab3 just prior to the *SER3* transcribed region, which could then terminate *SRG1* transcription to prevent interference with *SER3* [60, 147]. SRG1 can also be terminated through a polyA-dependent pathway at a site downstream of the NNS termination site(s) [147]. ChIP-exo revealed that RNAPII occupancy also shifts 3'- in *rrp6Δ* cells with increased RNAPII occupancy in the *SER3* gene (Figure 23A). An increase in specific *SRG1* transcripts including a *SRG1:SER3* chimeric transcript has previously been observed in Nrd1-depletion *rrp6Δ* cells [147]. Our data suggests that defective termination of *SRG1* can occur in *rrp6Δ* cells even in the absence of Nrd1 disruption (Figure 23). These data clearly show that Rrp6 is required for NNS-dependent termination of regulatory non-coding RNAs that participate in gene expression attenuation.

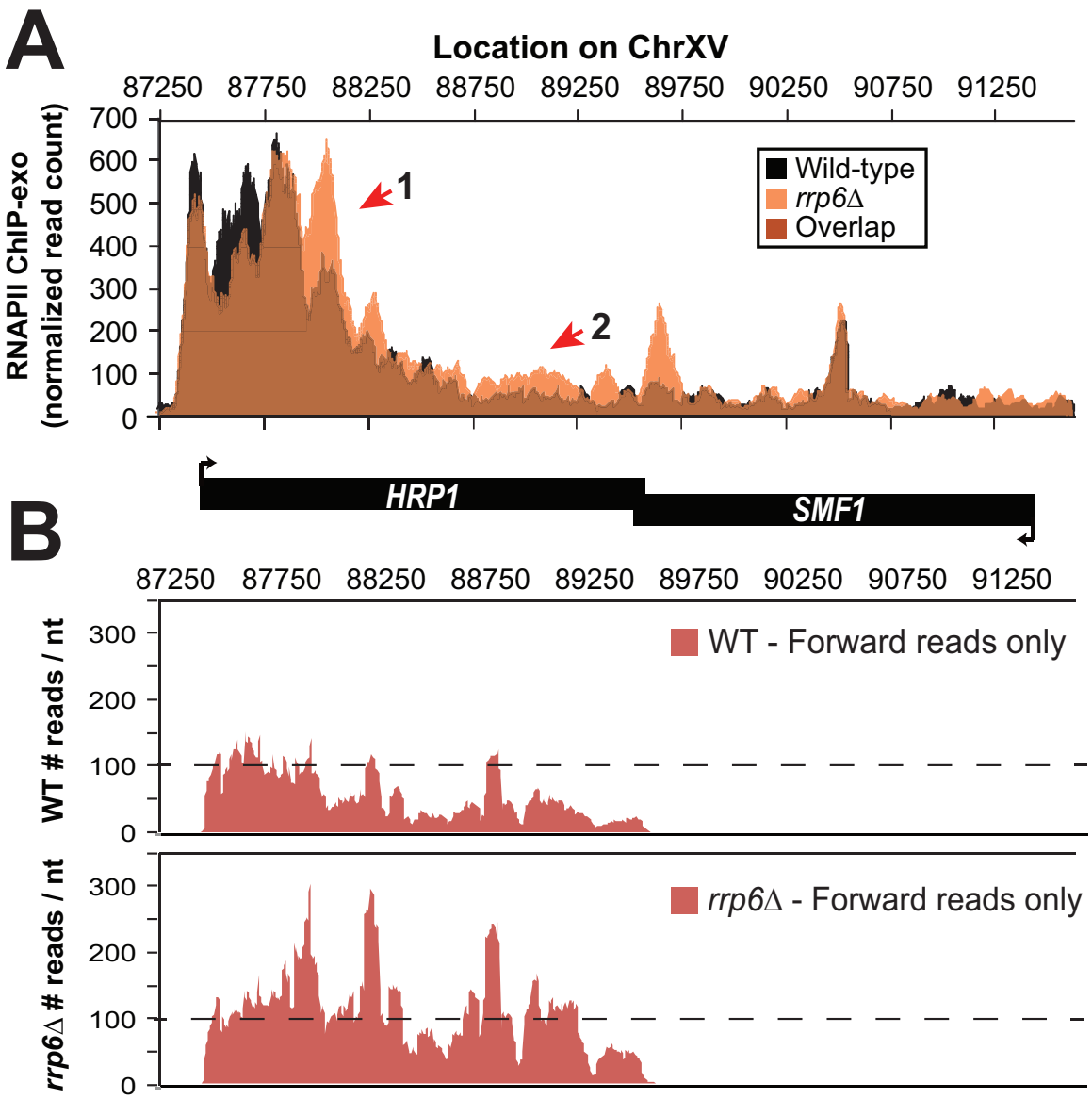

Figure 22: Efficient termination of RNAPII at the *HRP1* 5'-UTR requires Rrp6. (A) Rpb3-FLAG localization as determined by ChIP-exo sequencing reads mapped to the *HRP1* region. WT reads are in black, and *rrp6Δ* are in orange. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Length of mRNAs including untranslated regions are in black, CUTs are in gray, and NUTs are in aqua. (B) Graphical representation of strand-specific RNA-Seq reads mapped to *HRP1* transcribed region. Reads mapped to the positive strand are shown in red. No reads were mapped to the reverse strand at *HRP1*.

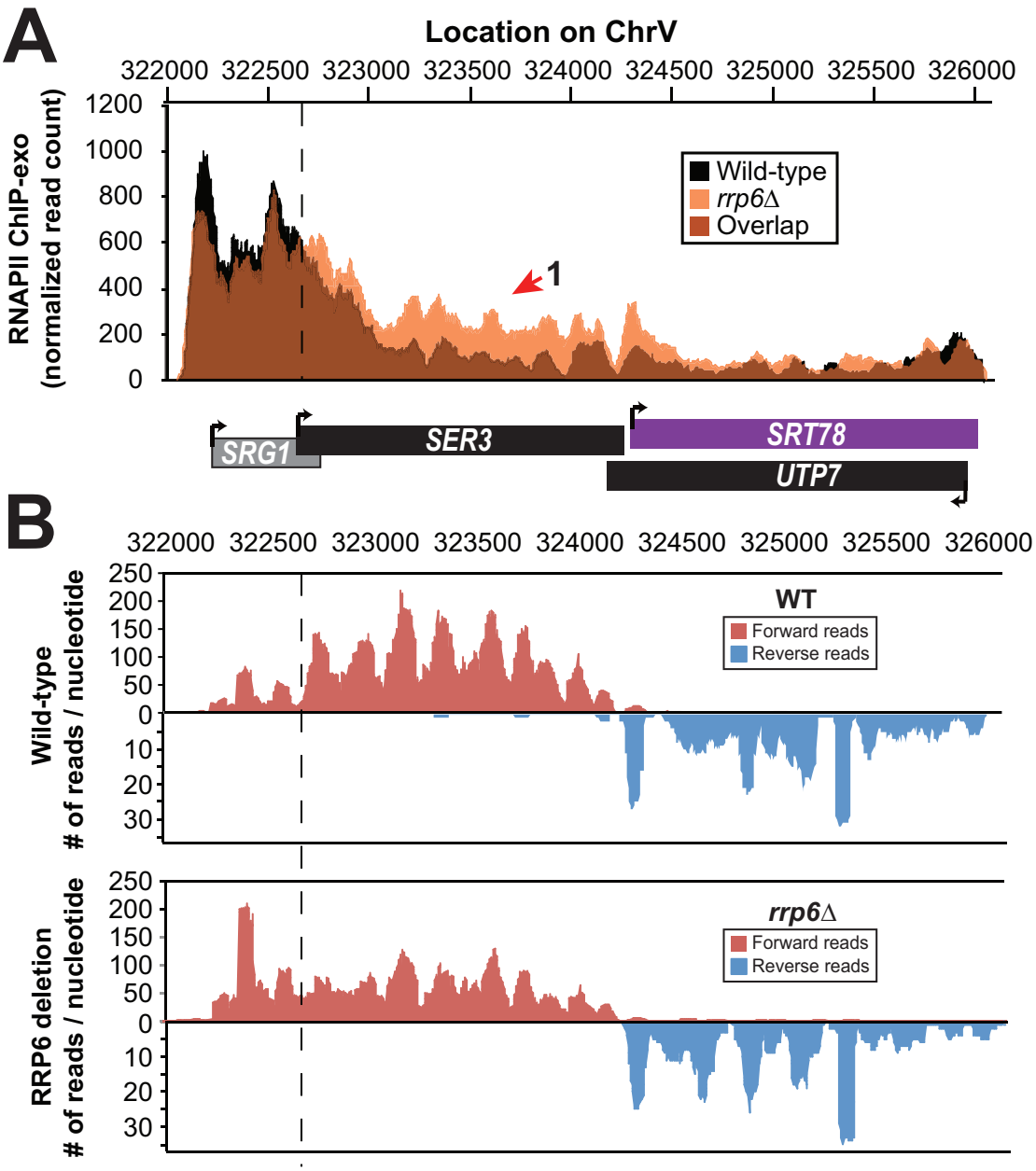

Figure 23: Rrp6 regulates RNAPII localization at *SRG1-SER3* independent of other NNS-pathway disruptions. (A) Rpb3-FLAG localization as determined by ChIP-exo sequencing reads mapped to the *SRG1-SER3* region. WT reads are in black, and *rrp6Δ* are in orange. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Length of mRNAs including untranslated regions are in black, CUTs are in gray, and NUTs are in aqua. Note that *SRG1* and *SER3* annotations are overlapping. (B) Graphical representation of strand-specific RNA-Seq reads mapped to *SRG1-SER3* transcribed region. Reads mapped to the positive-strand are shown in red while reverse reads are shown in blue.

There are multiple published examples of transcription interference by CUTs in which an antisense CUT regulates the expression of the sense gene in *cis* through transcript extension across the sense gene promoter [217, 226-229]. These lncRNA-type transcripts have also been shown to be required for efficient gene activation [230]. Expression of the *FMP40* transcript has been previously shown by several groups to be regulated by an NNS-terminated antisense transcript initiating at the 3' end of *FMP40* [41, 50, 231]. The *FMP40* antisense transcript (also known as CUT882 and *YPL222C-A)* is readily detectable in our RNA-Seq data from WT cells and it is significantly up-regulated in *rrp6Δ* cells as has been previously reported (Figure 24A, C) [72, 73, 232, 233]. Mapped sequencing reads suggest that the antisense transcript was significantly longer in *rrp6Δ* cells than in WT (specifically in the region of CUT882), supporting the hypothesis that NNS-dependent termination is not as efficient in *rrp6Δ* mutants. To determine the length of the antisense transcript in *rrp6Δ* cells and compare to an *nrd1*-ts mutant, we performed northern blot analysis with strand specific oligonucleotide probes to *FMP40* and its antisense transcript *YPL222C-A* (Figure 24D). *YPL222C-A* transcripts were detected as smears extending into the region of the CUT882 annotation suggesting that these transcripts terminate at multiple 3'-end locations as has previously been shown for other specific CUTs including *NEL025c* [71]. The *YPL222C-A* antisense transcripts were not detected in WT cells by northern blot even after an extended exposure time (Figure 24D). In the *nrd1*-ts mutant, there is an accumulation of a strong band ~4500 nt long in the *YPL222C-A* blot, and the *FMP40* transcript is undetectable as a consequence of

transcription interference. There is a band of the same size in the *rrp6Δ*, although it is not as abundant as that observed in the Nrd1 mutant. There are also shorter transcripts present in *YPL222C-A* blots derived from the *nrd1-ts* mutant that are also present in the *rrp6Δ* cells (Figure 24D). RNAPII localization in the *FMP40* region was determined by ChIP-exo and clearly shows that the majority of RNAPII was localized at the 5'-end of the antisense *YPL222C-A* transcript, even in WT cells (Figure 24B). The highest peaks of RNAPII localization at *YPL222C-A* are of similar intensity in WT and *rrp6Δ* cells, but RNAPII spreads 3'- in *rrp6Δ* cells and continues to be higher through the CUT882 annotation and past the *FMP40* promoter. These data indicate that the antisense transcript is terminated less efficiently in *rrp6Δ* cells leading to increased RNAPII occupancy downstream of the normal *YPL222C-A*/CUT882 termination site. The RNAPII occupancy data also suggest that CUT882 is an extended transcript of *YPL222C-A* that occurs as a consequence of inefficient RNAPII termination in *rrp6Δ* cells (Figure 24B, red arrows). In *nrd1Δ151-214* temperature sensitive mutants, *YPL222C-A* termination rarely occurs, resulting in a 4500 nt antisense transcript that is also seen at low levels in *rrp6Δ* strains (Figure 24D).

3. Rtr1 is a negative regulator of the NNS transcription termination pathway

3.1. RNA expression Data – Global analysis

To determine if a specific class of RNAPII transcripts were altered upon loss of the RNAPII CTD phosphatase *RTR1*, we performed RNA-Seq analysis of *rtr1Δ* cell total RNA and performed a Fisher's Exact Test on the

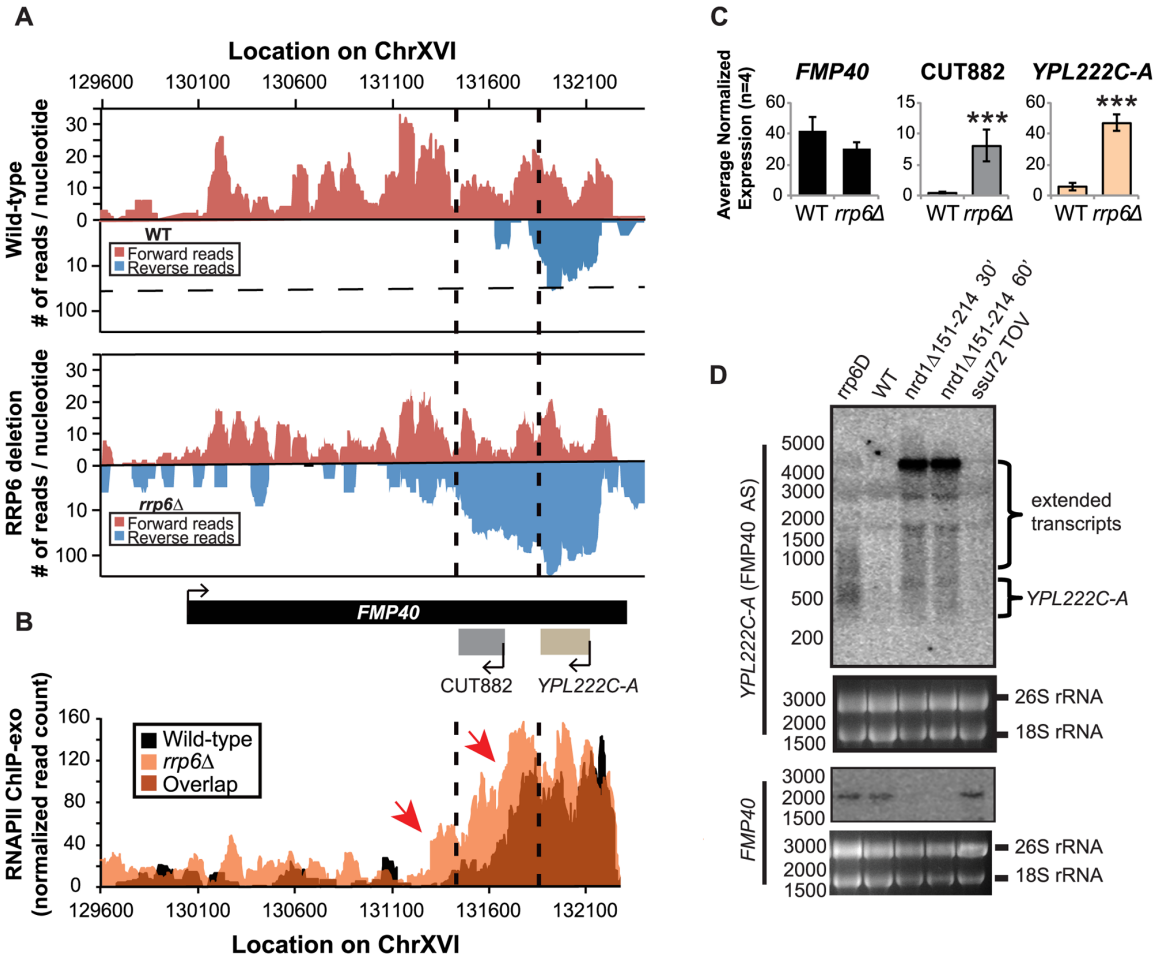

Figure 24: The *FMP40* antisense transcript *YPL222C-A* is extended in *rrp6Δ* cells as a result of inefficient RNAPII termination. (A) Graphical representation of strand-specific RNA-Seq reads mapped to *FMP40* region. Reads mapped to the positive-strand are on the top portion of the panel in red, whereas reads mapped to the negative-strand are on the bottom in blue. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Length of mRNAs including untranslated regions are in black, CUTs are in gray, dubious ORFs (an antisense transcript named *YPL222C-A*) are in tan, and bent arrows indicate direction of the TSS. The dotted black lines mark the 3' end of the annotations for CUT882 and *YPL222C-A*. (B) Rpb3-FLAG localization as determined by ChIP-exo sequencing reads mapped to the same region and aligned to (A). WT reads are in black, and *rrp6Δ* are in orange. Red arrows note areas of interest. (C) Average normalized read counts ± standard deviations for significantly altered transcripts in *rrp6Δ* cells versus WT (n=4). Three stars indicate a p-value of <0.001 as determined by an unpaired, two-tailed student's t-test. The colors of the bars in each graph correspond to the color representing the related annotation. (D) Strand-specific northern blot analysis using a 5' end labeled probes specific to either *FMP40* or *YPL222C-A* directly comparing *rrp6Δ* to mutants known to be defective in NNS termination.

differentially expressed transcripts identified by EdgeR analysis (Table 9, Figure 25). We analyzed transcripts that were both significantly increased and decreased in *rtr1Δ* cells compared to levels the total number of transcripts in each class in WT. A highly significant ($p<0.001$) number of snRNAs and open reading frame transcripts (ORF-Ts) are increased in *rtr1Δ* cells, and a significant number of antisense transcripts (ASTs) ($p<0.01$) and cryptic unannotated transcripts (CUTs) ($p<0.05$) were also up-regulated. More strikingly, multiple classes had a significant number of transcripts with decreased expression. ASTs, Nrd1 unterminated transcripts (NUTs), Ssu72-restricted transcripts (SRTs), and ORF-Ts all have a p-values of <0.001, and snRNAs are down-regulated with a significance of $p<0.01$. Many antisense annotations overlap with NUT annotations (example in Figures 29), indicating that they require Nrd1 for proper termination. Likewise, RNAPII-transcribed snRNAs are terminated by the NNS pathway, and many of the down-regulated ORF-Ts overlap with a NUT to some extent. This suggests that Rtr1 may have an important role in proper regulation of the Nrd1 termination pathway. SRTs are transcripts that were accumulated after loss of Ssu72 activity. Surprisingly, our data clearly show that those transcripts that are increased after loss of one Ser5 phosphatase (Ssu72) are often inhibited after the loss of the other Ser5 phosphatase (Rtr1), suggesting that the two enzymes have unique effects on the transcription cycle.

3.2. Immunoprecipitation of DNA sequences by Nrd1-bound RNAPII complexes

To determine the localization of TAP-tagged Nrd1 across an actively transcribed gene, we performed chromatin immunoprecipitation followed by exonuclease digestion and genome-wide sequencing (ChIP-exo) as described previously [211]. As noted earlier, Nrd1 does not bind DNA directly, rather it binds nascent RNA in a sequence-specific manner via its RNA RRM domain, and it binds RNAPII CTD repeats at Ser5-P via its CID. Therefore, this method detects regions of DNA bound by RNAPII that is also in complex with Nrd1 (See Figure 26A).

To confirm that the binding pattern was specific to Nrd1, we compared the Nrd1-TAP ChIP-exo reads to that of Rpb3-FLAG (Figure 26B, C). At the well-characterized protein-coding gene *PMA1*, RNAPII occupancy is relatively consistent across the entire length of the gene (Figure 26B). Nrd1-binding, in contrast, peaks in the 5'-end of the gene, approximately 500 nt past the transcription start site. Nrd1 has been reported to only function within <1000 nt from the transcription start site [44] and these data are consistent with those findings. Although there is a peak in the Rpb3-FLAG signal at the 5'-end, levels of bound RNAPII stay high across the entire length of the gene, while the amount of Nrd1 drops off quickly after the peak associated with the 5'end of the gene. Therefore, Nrd1 may play a role in RNAPII promoter-proximal pausing in yeast. Nrd1 has previously been reported to interact with the cap-binding complex [70]. Perhaps these interactions are required during a capping checkpoint that

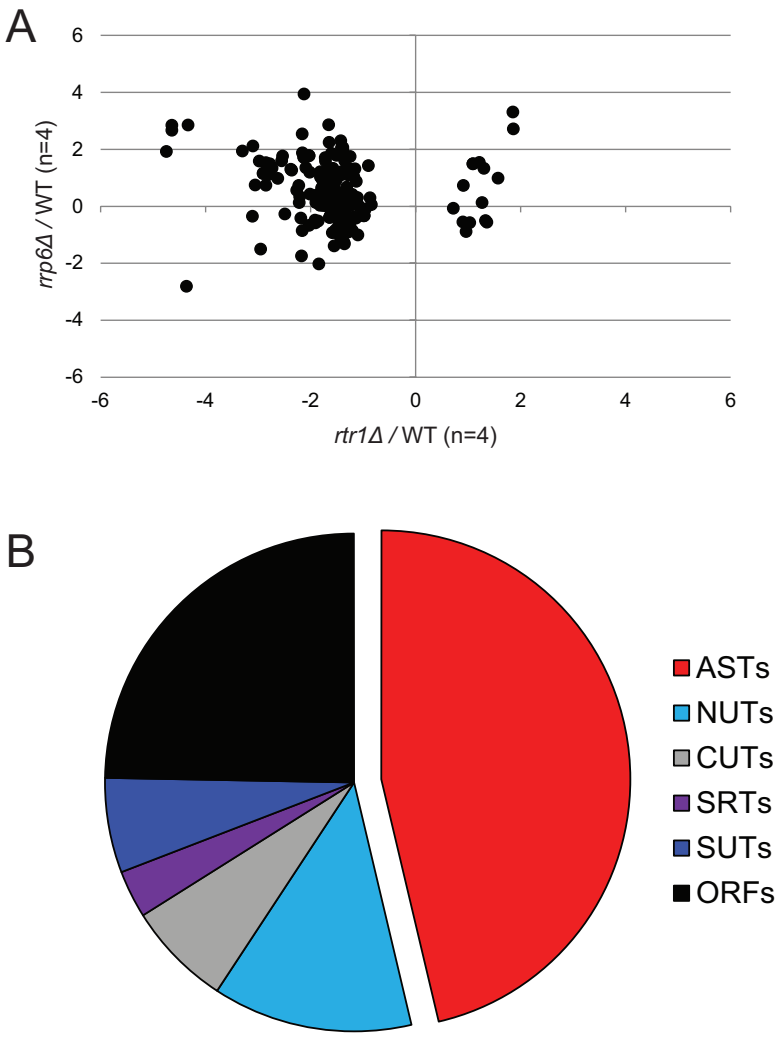

Figure 25: Characterization of differentially expressed transcripts in *rtr1Δ* cells. A) Expression plot for normalized RNA-Seq data for all transcripts differentially expressed in *rtr1Δ* cells. After sequencing reads were aligned to the yeast genome, reads mapped to annotated open reading frame transcripts (ORF-Ts), cryptic unstable transcripts (CUTs), stable unannotated transcripts (SUTs), and Nrd1 unterminated transcripts (NUTs) were used for differential expression analysis using edgeR. Log_2 of the fold-change values in *rrp6Δ* cells are plotted versus the log_2 of the fold-change values in *rtr1Δ* cells compared to WT across all biological replicates. B) Distribution of transcripts differentially expressed in *rtr1Δ* cells compared to WT across annotation categories. Colors correspond to colors used for these categories throughout this manuscript. For both (A) and (B) transcripts were included if $p \leq 0.005$ and FDR ≤ 0.32. A relaxed FDR was used with a stringent p-value cut-off because many of the cryptic transcripts are expressed at low levels and therefore have a large variance. The total number of transcripts at this cut-off is 194.

Category	Total	Increased in *rtr1Δ*	p-value (Fisher's Exact)	Decreased in *rtr1Δ*	p-value (Fisher's Exact)
AST	2576	32	0.0031	267	<0.0001
CUT	699	11	0.0231	39	0.458
MANUnit	346	0	0.1755	20	0.5315
NUT	1271	12	0.5687	104	<0.0001
snR	159	7	<0.0001	0	0.0069
SRT	180	3	0.3413	20	0.0002
SUT	675	3	0.4372	45	0.0384
ORFs	5742	22	<0.0001	78	<0.0001
Total	**11648**	**90**		**573**	

Table 9: Fisher's Exact Test to determine significance of dis-regulated transcript classes in *rtr1Δ* cells. The first column contains the names of the categories of RNAPII transcripts investigated. The total number of annotations included in our database for each category is in the second column. The number of transcripts significantly increased or decreased in *rtr1Δ* cells differential expression analysis is listed in the third and fifth columns, respectively. Transcripts were included if $p \leq 0.005$ and FDR ≤ 0.32. A relaxed FDR was used with a stringent p-value cut-off because many of the cryptic transcripts are expressed at low levels and therefore have a large variance. The total number of transcripts at this cut-off is 194. P-values calculated using the Fisher's Exact Test indicate the probability that the respective category of transcripts is significantly altered by the loss of *RTR1*.

ensures that mRNA is properly capped prior to RNAPII elongation. In the absence of proper capping, mRNA transcripts could be terminated during early transcription by the NNS pathway.

The *URA8*/*SOD1* locus illustrates the differences seen in the binding patterns of the Rpb3-FLAG and Nrd1-TAP strains when comparing transcripts with high (*URA8*) and low (*SOD1*) levels of consensus Nrd1-Nab3 RNA binding sites (Figure 26C). *URA8* encodes a CTP synthetase that carries out the ATP-dependent conversion of UTP to CTP. *URA8* is known to be regulated by alternative start site selection that is dependent on nucleotide availability [234]. There is a consensus Nrd1 binding site of TTTGTAAAGTT 40 nt upstream of the *URA8* ATG. The upstream, alternative start site is terminated by the NNS pathway in nutrient-rich conditions such as growth in YPD, resulting in low levels of full-length *URA8* transcription [152]. Our strains were grown in nutrient rich YPD media resulting in repression of *URA8* expression. Our ChIP-exo analysis of Rpb3-FLAG showed that the majority of RNAPII is localized at the 5'-end of the *URA8* gene (Figure 26C). This corresponds with the peak of Nrd1 binding at *URA8*, indicating that most of its transcription is being terminated by the NNS pathway, resulting in low-level transcription of full-length *URA8*. The *SOD1* gene is situated at the 3'-end of *URA8* and is transcribed in a convergent orientation. *SOD1* is not known to be regulated by NNS-dependent termination and does not contain Nrd1 RNA binding sites. Therefore, levels of Nrd1 association at the *SOD1* gene are much lower than at *URA8* even though total Rpb3-FLAG occupancy is higher at *SOD1* than at *URA8*.

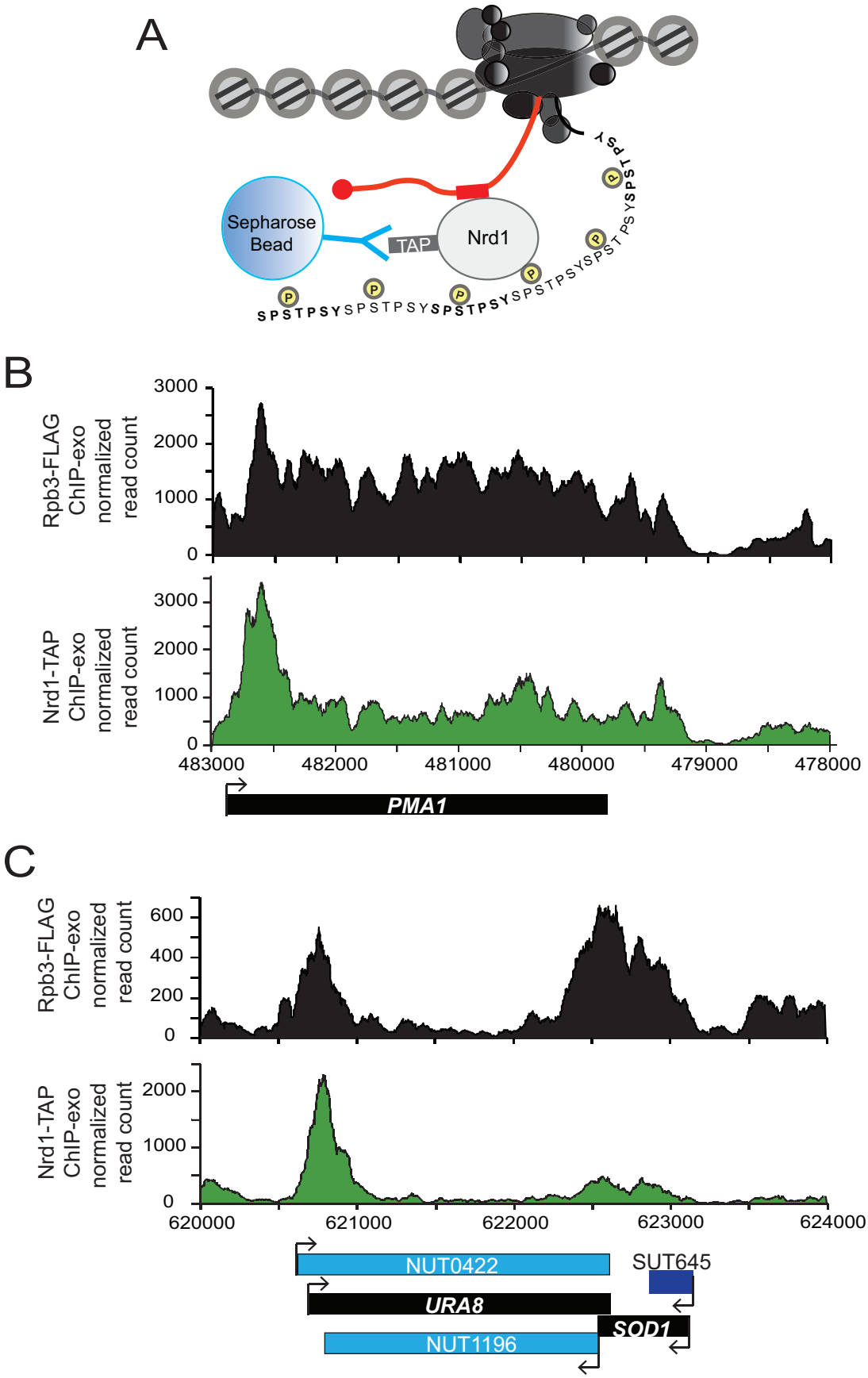
A
Sepharose Bead
TAP
Nrd1
SPSTPSYSPSTPSYSPSTPSYSPSTPSYSPSTPSYSPSTPSY
B
Rpb3-FLAG ChIP-exo normalized read count
3000
2000
1000
0
Nrd1-TAP ChIP-exo normalized read count
3000
2000
1000
0
483000
482000
481000
480000
479000
478000
PMA1
C
Rpb3-FLAG ChIP-exo normalized read count
600
400
200
0
Nrd1-TAP ChIP-exo normalized read count
2000
1000
0
620000
621000
622000
623000
624000
NUT0422
SUT645
URA8
SOD1
NUT1196

Figure 26: ChIP-exo analysis of the RNAPII- and RNA-binding protein Nrd1. A) Simplified schematic of immunoprecipitation of Nrd1-TAP by IgG-sepharose beads. Nrd1 binds to sequences in the RNA (red rectangle) via its RNA Recognition Motif (RRM domain). The RNA consensus sequences for Nrd1 binding are UGUAG and UGUAA. Nrd1 binds RNAPII at Ser5-P of the CTD of RNAPII via the Nrd1 CID. Therefore, DNA sequences detected by Nrd1-TAP ChIP represent regions of DNA bound by RNAPII bound to Nrd1. B and C) Graphical representation of Rpb3-FLAG (black) and Nrd1-TAP (green) occupancy in the WT strain as determined by ChIP-exo sequencing reads mapped to *PMA1* (B) and *URA8-SOD1* (C). The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Protein-coding mRNAs are in black, NUTs are in aqua, SUTs are in blue, and bent arrows indicate direction of the TSS.

To determine the locations of highest Nrd1-binding as measured by ChIP-Exo, we performed peak-calling analysis using the MACS2 algorithm [212]. Unsurprisingly, the top 100 ranking peaks were generally located at snRNAs and highly abundant, known Nrd1-regulated transcripts (Table 10). As described above, our ChIP-exo data identifies regions in the DNA where Nrd1-bound RNAPII complexes are localized. Nrd1 does not bind the DNA directly, but instead binds the nascent RNA. Transcriptome-wide binding sites of Nrd1 to RNA have previously been described using PAR-CLIP [50]. Again, the majority of the Nrd1 RNA binding sites are located near snRNAs and highly abundant, known Nrd1-regulated transcripts. We compared our top 100 Nrd1 DNA binding sites as measured by ChIP-exo to the published top 100 Nrd1 RNA binding sites as measured by PAR-CLIP.

From these two data sets, 44 binding sites occur in both lists. Of these 44 transcripts, 31 of them are located near snRNAs. This is not surprising because snRNAs are some of the most highly transcribed NNS-targeted transcripts and would therefore be expected to have the highest occupancy of RNAPII and Nrd1. Other peak locations occurring in both lists include genes with strong NNS terminators previously discussed such as *NRD1*, *URA2*, *URA8*, and *IMD2*. Interestingly, the ribosomal protein-coding genes *RPS14B* and *RPS9A* occur in both lists, as does the non-coding region of the rDNA repeat, *NTS2-1* (Table 10). *NTS2-1* is particularly surprising because the rRNA repeat region is transcribed by RNAPI and *NTS2-1* contains a known RNAPI termination site [235]. What function Nrd1 may be serving in this location is yet to be determined.

Of the top 100 Nrd1 DNA binding sites and top 100 RNA binding sites, 66 sites are reported independently in each list (Table 10). Creamer *et al*. also reported that Nrd1 PAR-CLIP and ChIP sites do not always overlap [50]. Specifically, the authors reported that 10 ribosomal protein coding genes were detected in their top 100 Nrd1 ChIP binding sites while they were not detected in their top 100 PAR-CLIP sites [50]. It is of note that a lower number of genomic locations that are present in ChIP-exo peak calling but not in the PAR-CLIP data are near snRNAs. Of these sites detected by one technique but not the other, 27/66 are near snRNAs in the ChIP data, but only 2/66 are near snRNAs in the PAR-CLIP data (Table 10).

One explanation for the relative abundance of snRNAs in the ChIP data compared to the PAR-CLIP data is that the ChIP data is dependent on the localization of RNAPII, whereas the PAR-CLIP localization is independent of RNAPII. It is likely that the Nrd1-binding peaks adjacent to snRNAs are more readily detectable, and therefore ranked higher in our list, due to the abundance of RNAPII localization and subsequent increased proximity of Nrd1 to the DNA. Additionally, Creamer *et al.* hypothesized that Nrd1 is able to bind to elongating RNAPII independent of RNA binding, allowing Nrd1 to bind Nrd1-binding sites in the nascent RNA as soon as it is available [50]. Finally, it is possible that these additional regions contain Nab3 RNA binding sites and that Nrd1 is recruited through its interaction with Nab3 not through direct RNA binding that would be detected by PAR-CLIP.

Gene	Chr.	Start position	End position	Summit relative to start	Fold-change	-log10 p-value	In both ChIP & CLIP?
snR8	XV	832350	832866	236	62.62	13301.30	yes
RPL9B	XIV	499664	500794	985	60.05	18900.26	
snR39B	VII	366209	366605	64	56.44	13540.35	yes
snR81	XV	234197	234845	384	48.76	6331.58	
NTS2-1	XII	459732	460624	786	45.02	20224.90	yes
snr72	XIII	297092	298824	1613	43.36	11983.66	
snr85	XIII	67343	67960	366	37.46	3005.84	yes
snr66	XIV	585650	586686	631	36.62	6909.90	
snR13	IV	1402990	1403459	144	34.53	10683.54	yes
snr64	XI	38711	39350	316	34.34	5147.62	
snr45	XVI	821693	822342	391	33.06	4957.72	yes
snr51	XVI	718291	719378	229	32.96	4743.62	
snr47	IV	541132	541864	328	32.78	6507.77	
snr48	VII	609443	610016	427	32.73	5303.05	
snr30	XII	198799	199759	729	32.24	7947.24	
IMD2	VIII	554099	554661	291	31.94	1843.78	
snR56	II	87983	88565	405	31.15	6096.91	yes
snR37	X	227697	228563	309	30.53	5580.30	
snR10	VII	345917	346778	451	30.08	6994.02	yes
PPS1	II	759668	759898	180	29.24	1609.37	
snr128	X	139119	140160	370	28.98	9025.75	yes
snr34	XII	899157	899858	348	28.75	3447.50	yes
URA2	X	172329	173132	555	28.21	3042.65	yes
snR83	XIII	626277	627042	523	28.00	3442.29	yes
IMD3	XII	1002267	1002952	276	27.63	3583.40	yes
snR161	II	306464	307340	643	27.58	2335.93	
snR40	XIV	89016	89715	405	27.44	8476.20	yes
snR71	VIII	411093	411911	350	27.42	3994.06	
snR11-CMC4	XIII	652175	653046	581	27.26	5004.97	yes
snR5	XV	842397	842984	366	26.33	4124.14	yes
snr17A	XV	780014	781013	731	26.18	4738.10	
snR80	V	51681	52409	399	25.95	3662.38	
snR79	XII	347918	348681	301	25.81	2773.99	
snR61	XII	794034	795148	366	24.45	4285.89	
snr4	V	424637	425457	479	24.19	3513.68	yes
RPS9A	XVI	404935	406116	556	24.15	3734.95	yes
snr50	XV	259321	260024	477	24.01	2181.64	yes
snr86	XIII	761640	762471	358	23.76	3336.11	yes
TYE7	XV	977246	978350	809	23.71	2422.06	
snR60	X	348662	349402	410	23.67	3159.60	yes
snR37	X	227697	228563	255	23.51	3857.54	

Gene	Chr.	Start position	End position	Summit relative to start	Fold-change	-log10 p-value	In both ChIP and CLIP?
snR10	VII	345917	346778	451	30.08	6994.02	yes
PPS1	II	759668	759898	180	29.24	1609.37	
snr128	X	139119	140160	370	28.98	9025.75	yes
snr34	XII	899157	899858	348	28.75	3447.50	yes
URA2	X	172329	173132	555	28.21	3042.65	yes
snR83	XIII	626277	627042	523	28.00	3442.29	yes
IMD3	XII	1002267	1002952	276	27.63	3583.40	yes
snR161	II	306464	307340	643	27.58	2335.93	
snR40	XIV	89016	89715	405	27.44	8476.20	yes
snR71	VIII	411093	411911	350	27.42	3994.06	
snR11-CMC4	XIII	652175	653046	581	27.26	5004.97	yes
snR5	XV	842397	842984	366	26.33	4124.14	yes
snr17A	XV	780014	781013	731	26.18	4738.10	
snR80	V	51681	52409	399	25.95	3662.38	
snR79	XII	347918	348681	301	25.81	2773.99	
snR61	XII	794034	795148	366	24.45	4285.89	
snr4	V	424637	425457	479	24.19	3513.68	yes
RPS9A	XVI	404935	406116	556	24.15	3734.95	yes
snr50	XV	259321	260024	477	24.01	2181.64	yes
snr86	XIII	761640	762471	358	23.76	3336.11	yes
TYE7	XV	977246	978350	809	23.71	2422.06	
snR60	X	348662	349402	410	23.67	3159.60	yes
snR37	X	227697	228563	255	23.51	3857.54	
snR63	IV	322188	323650	904	23.49	2597.43	
snR32	VIII	381569	382093	269	23.38	4229.31	
snR35	XV	758851	759507	423	23.29	3688.46	yes
snr189-snR65	III	177052	178757	1463	23.25	2731.82	yes
RPL41A	IV	129961	130561	164	23.24	2444.40	
snR79	XII	347918	348681	386	23.06	2362.00	
RPL7B-snR59	XVI	173073	174138	559	22.89	3429.21	yes
TDH3	VII	882423	883959	195	22.41	3921.99	yes
snR58	XV	135689	136324	265	21.61	1588.59	
RPL25	XV	80363	82938	1027	21.59	2961.70	
snr37	X	227697	228563	195	21.50	3393.44	
snr4	V	424637	425457	320	21.48	2959.42	yes
GIT1	III	297067	297226	99	21.40	1032.56	
MCM21	IV	1103799	1104154	282	20.99	1004.52	
NRD1	XIV	173699	174663	656	20.89	2507.13	yes

Gene	Chr.	Start position	End position	Summit relative to start	Fold-change	-log10 p-value	In both ChIP and CLIP?
snR49	XIV	716069	716708	404	20.84	1833.15	yes
snr3	X	663669	664742	354	20.48	2166.19	yes
NUP133	XI	594093	594258	108	19.60	1337.37	
HTZ1	XV	303315	303586	145	19.20	1200.63	
MTO1	VII	55514	56171	418	18.85	1532.38	yes
SRG1	V	322136	323577	535	18.84	2429.55	
URA8	X	620608	621119	181	18.66	1219.66	yes
snr35	XV	758851	759507	330	18.60	2659.08	yes
snR18	I	142115	143699	1347	17.85	1268.30	
snr82	VII	316816	317485	411	17.83	1176.22	yes
tG(UCC)O	XV	110105	111113	757	17.41	1132.50	
snR7-L	VII	938983	939738	331	17.40	781.23	
snr9	XV	407564	408218	285	17.37	2511.44	
HOS4	IX	153875	154124	117	17.21	750.58	
snr43	III	107016	107743	363	17.21	1329.34	
ydl241w	IV	20959	21223	56	17.19	749.34	
ylr257w	XII	657300	658083	479	17.14	745.61	
YBR074w	II	387692	387851	96	17.00	736.92	
GNP1	IV	1467714	1468578	633	16.76	1039.31	
snr33	III	142027	142513	170	16.54	1685.55	
ADE12	XIV	234249	234828	104	16.52	1740.35	
ADH1	XV	159176	160268	227	16.28	1220.50	
snr46	VII	545150	546099	459	16.19	1299.50	yes
snR72-78	XIII	297092	298824	1530	16.17	2957.56	
SHM2	XII	258774	259511	85	16.12	1713.66	
tS(CGA)C	III	227847	228577	249	15.98	770.66	
snr42	XI	558566	559361	319	15.98	1560.63	yes
RPS2	VII	277461	278758	1112	15.92	1440.35	yes
tA(UGC)A	I	166122	166363	119	15.75	657.44	
RPS14B	X	73755	74982	469	15.71	2055.64	yes
SWI3	X	92270	92421	55	15.52	786.49	
snr53	V	61061	62026	771	15.14	1414.02	
intergenic_2	XIV	784009	784333	251	15.01	610.81	
TOD6	II	117864	118544	477	14.87	602.53	
GRS1	XVI	701917	702486	207	14.76	614.67	
intergenic_4	II	9429	9905	48	14.66	589.57	
intergenic_5	VIII	562435	562643	94	14.55	582.53	
snr69	XI	364727	365507	267	14.52	804.83	
IRC2	IV	679625	679962	159	14.39	573.18	
OGG1	XIII	152850	153067	52	14.31	568.51	
BG_YJR107	X	628105	628688	86	14.12	556.90	

Gene	Chr.	Start position	End position	Summit relative to start	Fold-change	-log10 p-value	In both ChIP and CLIP?
CCW12	XII	369295	370228	646	14.01	1684.52	yes
PCF11	IV	923095	923923	94	14.01	549.96	

Table 10: Top 100 peaks in Nrd1 ChIP-Exo data. Chromosomal location of Nrd1 ChIP-exo peaks are listed with the name of the closest annotation gene, in order of descending fold-change values. The highest point in the peak is listed as "Summit relative to start." Peaks were considered included in both ChIP and PAR-CLIP top 100 site lists if the chromosomal location overlapped. ChIP peaks are much broader than PAR-CLIP peaks, and in some instances, two PAR-CLIP peaks were included in one ChIP peak. These double peaks were counted as two locations detected in both data sets for Figure 27.

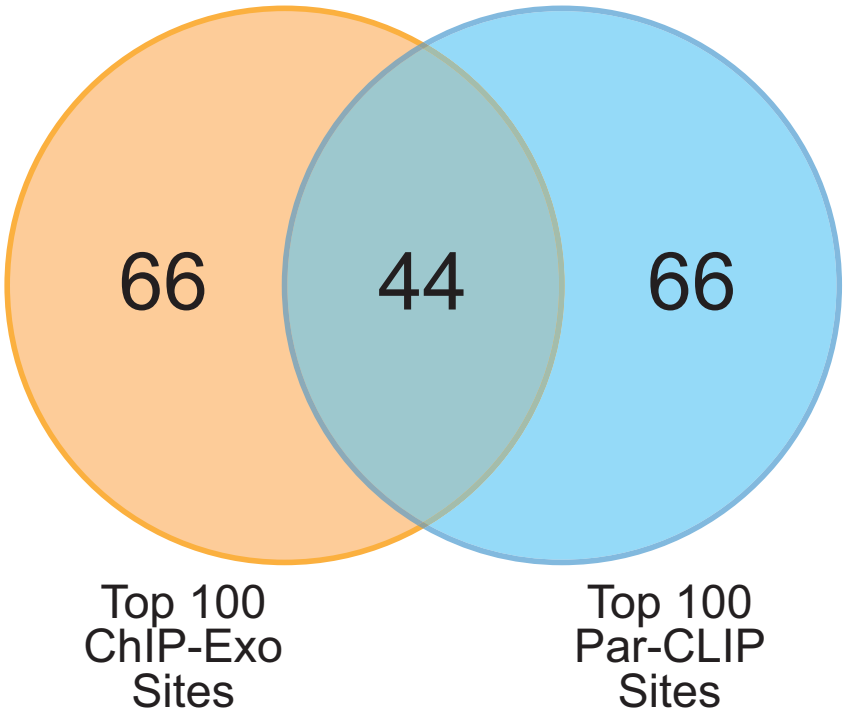

Figure 27: Comparison of top Nrd1-binding sites as detected by ChIP-exo versus PAR-CLIP. The DNA footprint of Nrd1-bound RNAPII was measured by Nrd1-TAP ChIP-exo. The direct RNA-binding sites of Nrd1 were measured by Creamer *et al.* [50] by Nrd1-His PAR-CLIP. The top 100 sites identified in each data set were compared. A total of 44 binding sites were identified in both lists.

3.3. *RTR1* deletion increases the recruitment of Nrd1 to the Ser5-P CTD.

To understand the effects on NNS-dependent termination in the *RTR1* deletion strain, we wanted to determine if the ratio of bound Nrd1 to RNAPII was increased at genes with known Nrd1 binding sites. Using the normalized number of reads mapped per base pair for each ChIP-exo dataset, we divided the reads from Nrd1-TAP by Rpb3-FLAG in the corresponding genetic backgrounds. This resulted in a ratio of Nrd1-bound RNAPII to Rpb3 per base that is shown at two well-characterized Nrd1-regulated transcripts *URA2* (Figure 28A) and *NEL025C* (overlapping with the NUT0952 annotation, Figure 28B). Expression of *URA2* is repressed in the presence of uracil, and previous ChIP experiments have shown that RNAPII occupancy correlates with expression in the *URA2* ORF, but not at the promoter where it is always present [236]. The promoter region contains an upstream CUT that is transcribed in the presence of uracil and terminated by the NNS pathway [152]. Our data from yeast grown in YPD medium indicate that there is a high occupancy of Nrd1 at this upstream CUT in WT cells, in agreement with previous findings (CUT680, Figure 28A). Additionally, the ratio of Nrd1 to RNAPII bound to genes increases nearly two-fold in the absence of *RTR1* (Figure 28A, orange), suggesting that Rtr1 activity in WT cells may limit the co-occupancy of Nrd1 and RNAPII at this location.

NEL025C, the first characterized CUT, originates from a divergent promoter at the *DLD3* gene [71]. NEL025C is expressed as either a long form (about 620 nt, corresponding to the annotations for SUT503, CUT541, and CUT542 combined) or a series of heterogeneous shorter transcripts, all of which

were shown to be terminated by the NNS-dependent termination machinery and degraded by Rrp6 [71, 147]. Using ChIP-exo, we identified a large peak in the ratio of Nrd1 to RNAPII associated along the length of the NUT0952 annotation (Figure 28B). This peak appears to be primarily due to an absence of Rpb3 localization at this region in WT cells, which could indicate that RNAPII termination is more efficient in the absence of Rtr1. This peak also overlaps with the *NEL025C* gene, with the highest peak localized at the 3'-end of the CUT542 annotation. There is also a high ratio of Nrd1 to RNAPII observed in the WT strain in the middle of the TSSs for *DLD3* and NUT0952. In the *rtr1Δ* strain, the Nrd1/RNAPII ratio is increased from 4-6 fold in WT to up to 18 in *rtr1Δ* cells at the peak of mapped DNA reads, indicating that deletion of *RTR1* increases Nrd1 co-occupancy with RNAPII.

3.4. Altered Composition of Affinity Purified Transcription Elongation and Termination Complexes

Because we detected an increased ratio of Nrd1:RNAPII co-localization at specific NNS-target gene transcripts in *rtr1Δ* cells as judged by our ChIP-exo studies, we used affinity purification mass spectrometry analysis to determine if the Nrd1:RNAPII interaction was increased upon deletion of *RTR1*. Furthermore, because it has previously been shown that loss of *RTR1* results in 3'-end processing defects at some mRNAs [8], we wanted to determine how the deletion of *RTR1* affected the composition of proteins associated with RNAPII termination complexes. We performed affinity purification mass spectrometry

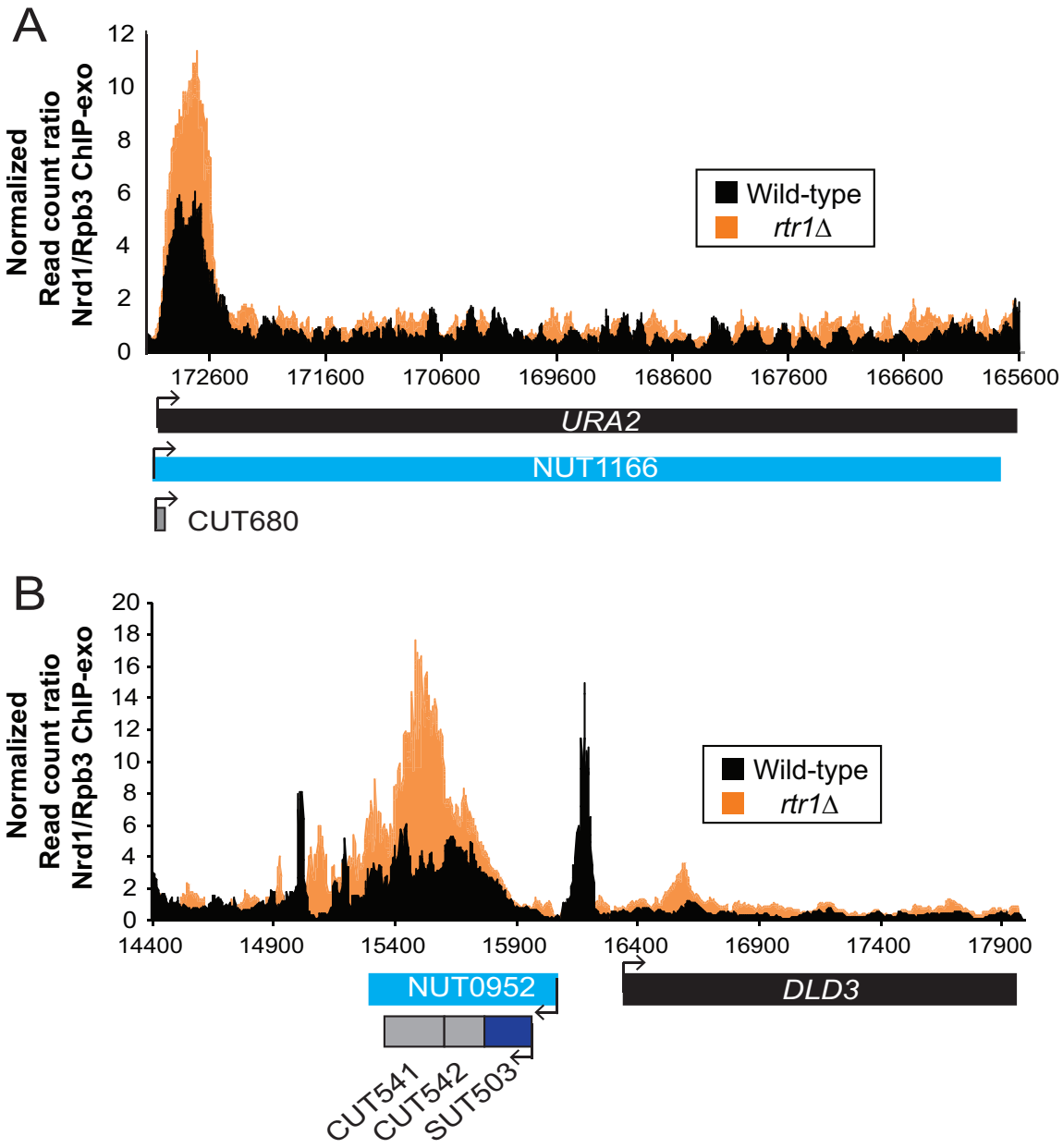

Figure 28: Co-occupancy of Nrd1/RNAPII is increased genome-wide with deletion of RTR1 as judged by Rpb3-FLAG and Nrd1-TAP ChIP-exo analysis. Graphical representation of the ratio of Nrd1 to Rbp3 read counts per base mapped to *URA2* (A) and *DLD3* region (B). Data derived from WT cells is in black, and those from the *rtr1Δ* strain is in orange. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs, each to scale. Protein-coding mRNAs are in black, NUTs are in aqua, SUTs are in blue, CUTs are in gray, and bent arrows indicate direction of the TSS. The transcript approximated by the NUT0952 annotation has also been referred to as NEL025C in the literature [71].

analysis of subunits of CPF, Rtt103/Rat1/Rai1, CF1a, and Nrd1 from WT and *rtr1Δ* strains as shown in Figure 29. The data was then analyzed by hierarchical clustering as previously described [237]. The resulting cluster separates the proteins identified by mass spectrometry into their respective termination complexes (Figure 29). Subunits that associate more transiently with the complexes are also detected but have a lower NSAF value than the core complex components.

Rpb3-TAP isolation from WT cells followed by mass spectrometry successfully identified all subunits of RNAPII. Regulators of RNAPII termination were not detected in the Rpb3-TAP isolation, likely because of the dynamic nature of recruitment of termination factors during the transcription cycle. Cdc73 is a member of the Paf complex (PAF-C), which is involved in RNAPII transcription elongation and has been implicated in transcription termination [238-242]. All subunits of PAF-C and RNAPII were detected in Cdc73-TAP purifications. Rtr1, an elongation specific interaction partner of RNAPII, was found to primarily interact with all of the subunits of RNAPII and PAF-C in WT cells. We also looked at the interactions of Ssu72, the other Ser5 phosphatase, in WT and *rtr1Δ* cells. Ssu72 is a member of the CPF complex, and all CPF subunits were detectable in WT and *rtr1Δ* cells with the exception of Yth1. Yth1 has previously been shown to interact with Ssu72-purified complexes at low levels [188]. Subunits of PAF-C were also detected in Ssu72-FLAG samples. Of interest, Ssu72-FLAG and presumably the CPF complex interact with Yra2, a protein involved in RNA export in WT, but this interaction in lost in the *rtr1Δ*

strain. We did not detect any increased protein-protein interactions in the Ssu72-FLAG *rtr1Δ* purification that would suggest a clear compensatory function of Ssu72 in the absence of *RTR1*.

Our data indicate that loss of *RTR1* disrupts the association of a number of subunits in the RNAPII transcription termination machinery. Many proteins that normally interact with Rpb1 via the CTD were disrupted upon loss of *RTR1* (green box). Rtt103 normally binds RNAPII through the Ser2-P mark on the CTD. In WT cells, we detect Rtt103 interaction with many of the RNAPII subunits and interactions with known binding partners Rat1 and Rai1. However, we did not detect Rat1 or Rai1 (red box) or any subunits of RNAPII in the Rtt103 purification in *rtr1Δ* cells. Likewise, Pcf11 also binds to Ser2-P modified RNAPII. In this cluster, we show two replicates of the Pcf11 purifications in both WT and *rtr1Δ* to illustrate the reproducibility of this method. Similar to the data from the Rtt103 purifications, Pcf11 interacts with Rpb1 (green box) and other subunits of RNAPII in the WT cells, but these interactions are not detectable in *rtr1Δ* cells. In addition, we observed decreases in the association of the CFIa subunit Clp1 in Pcf11-TAP *RTR1* deletion purifications. These findings are consistent with previous *in vitro* studies that found that Pcf11 and Rtt103 binding to Ser2-P modified CTD peptides was much stronger than to Ser5-P [32].

Interestingly, Nrd1-TAP appears to have increased interactions with both Rpb1 and the Cap Binding Complex (Sto1 and Cbc2) in purifications from *RTR1* deletion strains relative to WT (boxed in aqua in Figure 29). These data support our ChIP-exo data shown in Figure 28 that show an increased ratio of

Nrd1:RNAPII association at specific genes in the *rtr1Δ* strain. Nrd1 has been well documented to interact with the Cap Binding Complex [70], and, although it is not detectable in the WT dataset presented here. The interaction is likely to occur below the detectable threshold due to the low abundance of Nrd1 in WT cells. Of importance, Nrd1 and its interacting partner Nab3 are detected at similar levels in both WT and *rtr1Δ* strains, and the ratio of Sto1 and Cbc2 compared to Nrd1 bound to Nab3 is dramatically increased in the *RTR1* mutant. Interestingly, this increase of Sto1/Cbc2 interaction with Nrd1 is even greater than the increase of interaction between Rpb1 and Nrd1. The increased interaction between Nrd1 and the cap binding proteins may play a role in promoting the Nrd1/Nab3 interaction with the nascent RNA. These findings suggest that the loss of *RTR1* increases the affinity of the Nrd1-Nab3 complex for RNAPII *in vivo*, likely through increasing the number of Ser5-P modified CTD repeats. This suggests that regulation of Rtr1 activity and or expression level could be a key regulatory step for transcription termination through the NNS pathway.

3.5. Functional consequences of increased Nrd1 interaction with RNAPII in *RTR1* deletion cells.

Rtr1 appears to have a role in regulating antisense transcription at poorly expressed genes such as *YKL151C*, a known target of the NNS pathway [50]. RNA-Seq analysis revealed that the antisense transcript in the 3'-end of *YKL151C* appears shorter in *rtr1Δ* cells relative to WT. Annotated as "AS_Unit4881" (indicated by the red bar, Figures 30, 31), this antisense transcript

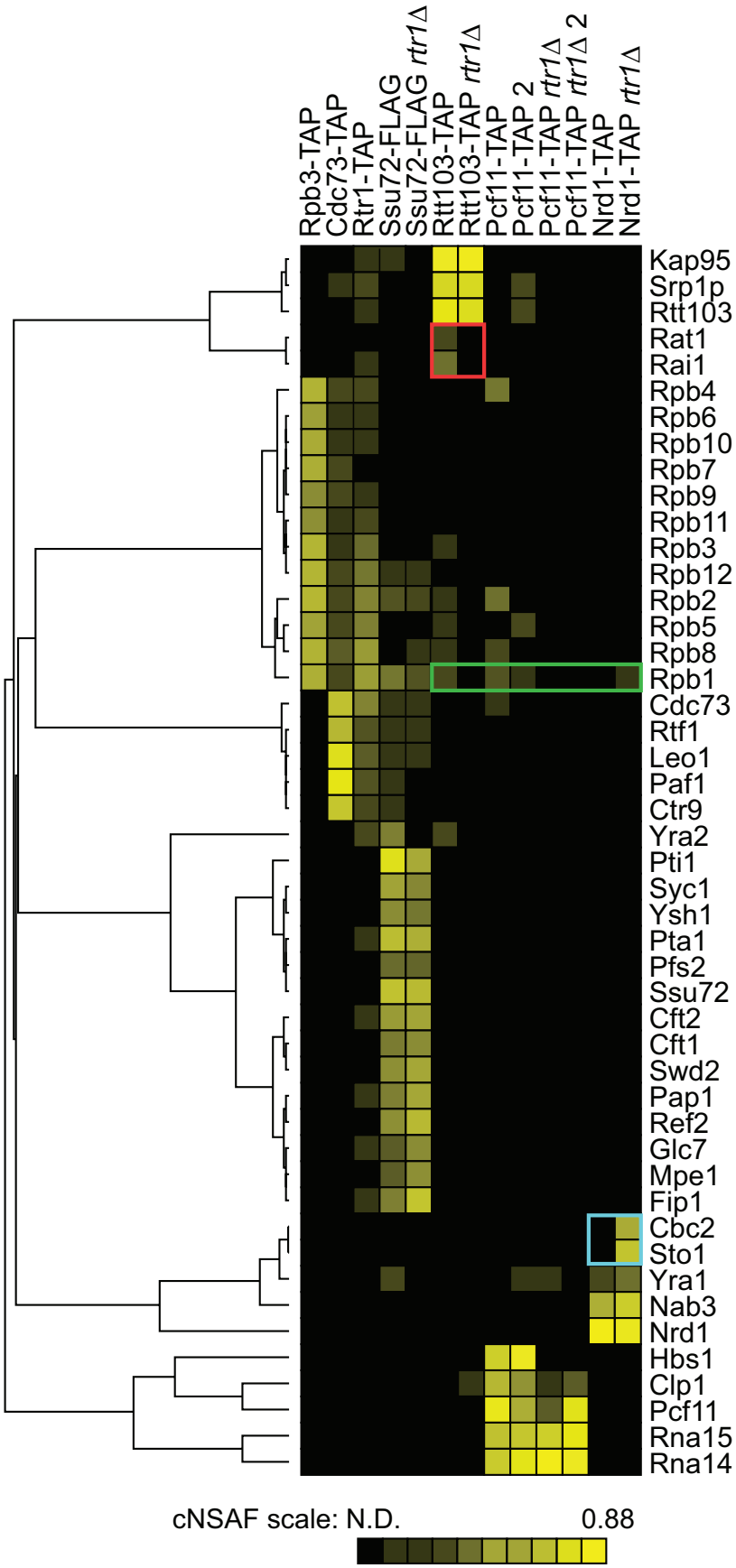
Rpb3-TAP
Cdc73-TAP
Rtr1-TAP
Ssu72-FLAG
Ssu72-FLAG rtr1Δ
Rtt103-TAP
Rtt103-TAP rtr1Δ
Pcf11-TAP
Pcf11-TAP 2
Pcf11-TAP rtr1Δ
Pcf11-TAP rtr1Δ 2
Nrd1-TAP
Nrd1-TAP rtr1Δ
Kap95
Srp1p
Rtt103
Rat1
Rai1
Rpb4
Rpb6
Rpb10
Rpb7
Rpb9
Rpb11
Rpb3
Rpb12
Rpb2
Rpb5
Rpb8
Rpb1
Cdc73
Rtf1
Leo1
Paf1
Ctr9
Yra2
Pti1
Syc1
Ysh1
Pta1
Pfs2
Ssu72
Cft2
Cft1
Swd2
Pap1
Ref2
Glc7
Mpe1
Fip1
Cbc2
Sto1
Yra1
Nab3
Nrd1
Hbs1
Clp1
Pcf11
Rna15
Rna14
cNSAF scale: N.D.
0.88

Figure 29: Quantitative proteomic analysis of complexes involved in RNAPII transcription termination in WT or *rtr1Δ* strains. The protein used for purification, type of C-terminal tag, and genotype are shown at the top of the figure. Two independent biological replicates are shown for Pcf11 purifications to show the reproducibility of protein identifications. Proteins identified and their respective complexes are shown to the right, and their NSAF values are indicated according to the scale on the bottom left. The dendogram to the left of the cluster represents the abundance relationship between the purified proteins as determined by clustering analysis performed as previously described [201, 202].

is detectable at low levels in WT and much more abundant in both strains lacking Rrp6. We created an *rtr1Δ rrp6Δ* double deletion strain to uncover transcripts affected by the loss of *RTR1* that would be degraded by Rrp6. However, we discovered that there is a requirement for Rrp6 for termination of specific NNS-target transcripts, as mentioned above [243]. Therefore, we were interested to determine what effects the interplay between Rrp6 and Rtr1 in the NNS-dependent termination pathway would have. In the *rtr1Δ rrp6Δ* double deletion strain, reads aligning to the AS_Unit4881 region to the left of the dashed line, where the transcript was detectable in the *rtr1Δ* deletion strain, are much more abundant than reads downstream of this site. This region was previously annotated as NUT0447, a transcript that requires Nrd1 for termination. Rpb3-FLAG occupancy is decreased in this region in *rtr1Δ* cells compared to WT cells, but Nrd1-TAP is increased as measured by ChIP-Exo (Figure 31). Therefore, the ratio of Nrd1 bound to RNAPII is higher in *rtr1Δ* cells as we have previously observed at other NNS target genes (Figure 30). Interestingly, the highest amount of change was observed in the 5'-end of the antisense transcript that could facilitate early termination in the absence of *RTR1*, in agreement with the RNA-Seq results (Figure 30).

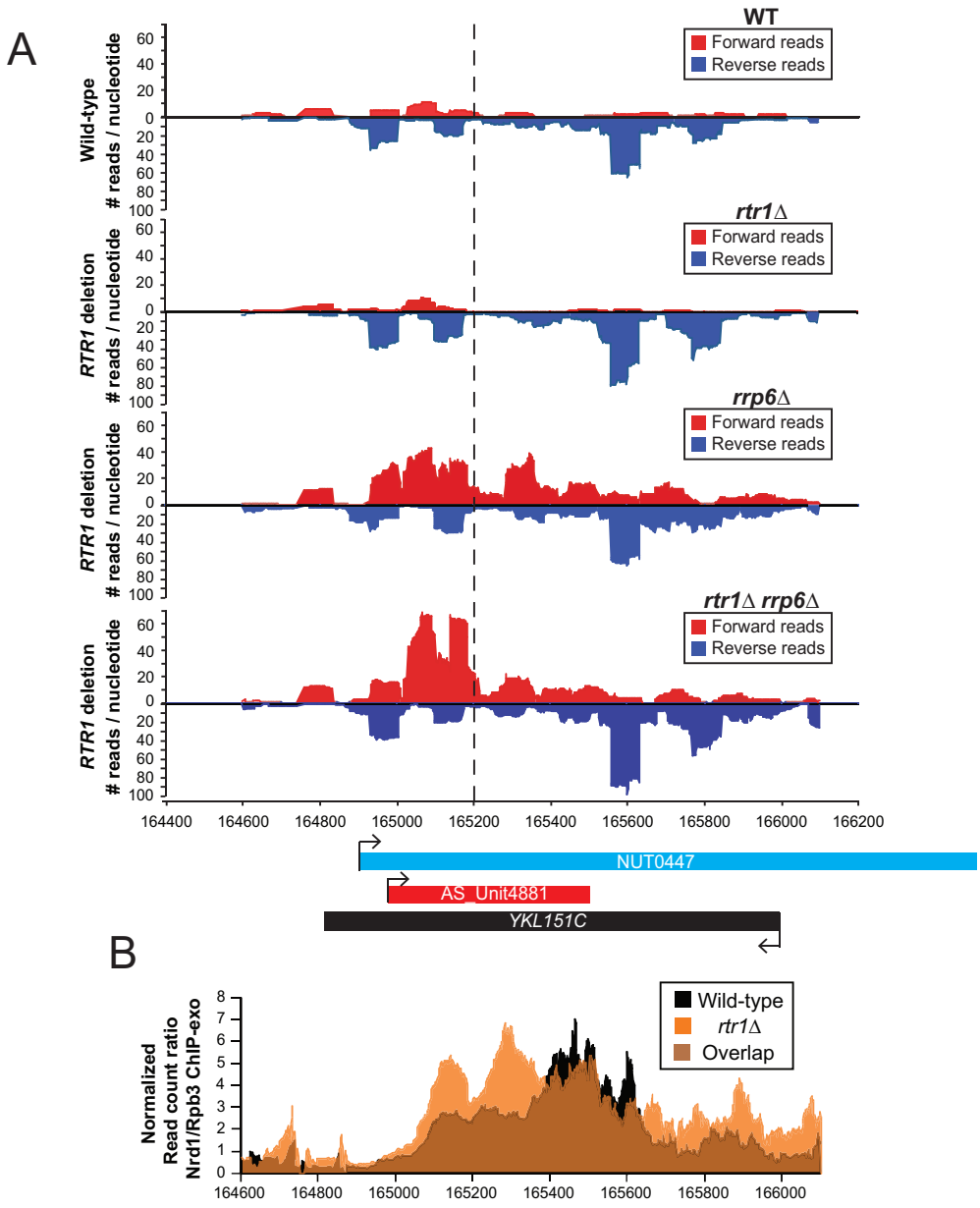

Figure 30: *RTR1* deletion shortens the NNS-terminated antisense transcript at *YKL151C*. (A) Graphical representation of strand-specific RNA-Seq reads mapped to *YKL151C* region. Reads mapped to the positive strand are on top in red, while reads mapped to the negative strand are on the bottom in blue. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Processed lengths of mRNAs are in black, NUTs are in aqua, antisense transcripts are in red, and bent arrows indicate direction of the TSS. (B) Graphical representation of the ratio of Nrd1 to Rbp3 read counts per base mapped to the same region as in (A) Data from WT is in back, *rtr1Δ* is in orange.

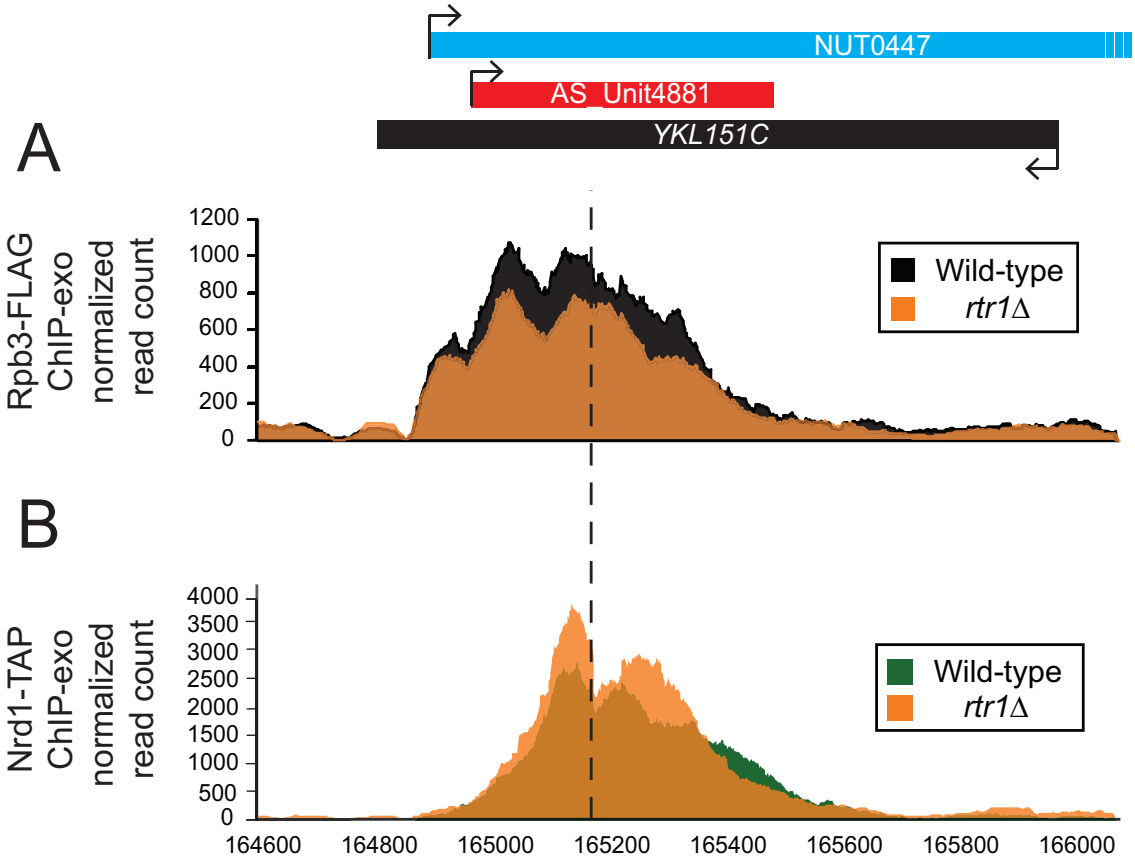

Figure 31: *RTR1* deletion alters Rpb3 and Nrd1 occupancy at *YKL151C*. Rpb3-FLAG (A) and Nrd1-TAP (B) localization as determined by ChIP-exo sequencing reads mapped to *YKL151C*. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Protein-coding mRNAs are in black, antisense transcripts are in red, and bent arrows indicate direction of the TSS. The dotted black line marks the estimated TTS of the antisense transcript in *rtr1Δ* cells. WT Rpb3 reads are in black, WT Nrd1 reads are in green and *rtr1Δ* reads from both strains are in orange.

It has been well-documented that *FMP40* is regulated by an NNS-terminated antisense transcript also known as CUT882 and *YPL222C-A* [41, 50, 231], and we have previously reported that termination of this RNA is Rrp6-dependent [243]. As determined by RNA sequencing and Rpb3-FLAG ChIP-exo, YPL222C-A is both more abundantly expressed and longer in *rrp6Δ* cells [243]. By the same methods, *rtr1Δ* appears to have the opposite effect (Figure 32). RNA expression of *YPL222C-A* is down regulated, and RNAPII localization is shifted 5'- relative to the antisense transcript compared to WT. Not only does RNAPII appear to travel a shorter distance during *YPL222C-A* transcription in *rtr1Δ* cells, but there is also an increased occupancy of RNAPII at the 5'-end of *YPL222C-A* that may be paused. Nrd1 localization is also increased in the *rtr1Δ* strain, and the first peak of Nrd1 binding aligns well with the RNAPII peak (Figure 32A). This suggests that loss of *RTR1* is promoting Nrd1 localization at this gene location, resulting in an increase of RNAPII that is paused in this region. In the *rtr1Δ rrp6Δ* strain, *YPL222C-A* expression is markedly higher than in WT cells and is similar to that observed in the *rrp6Δ* strain, which suggests that Rrp6 is required for the decrease in *YPL22C-A* expression in *RTR1* deleted cells. These findings support a model in which transcription elongation of *YPL222C-A* is positively regulated by the removal of Ser5-P by Rtr1. In the absence of Rtr1, Ser5-P levels remain high facilitating RNAPII termination via the NNS pathway. In WT cells, the removal of Ser5-P by Rtr1 decreases Nrd1-Nab3 binding and subsequent NNS pathway action at the earliest termination sites, resulting in synthesis of transcripts that are able to escape early termination.

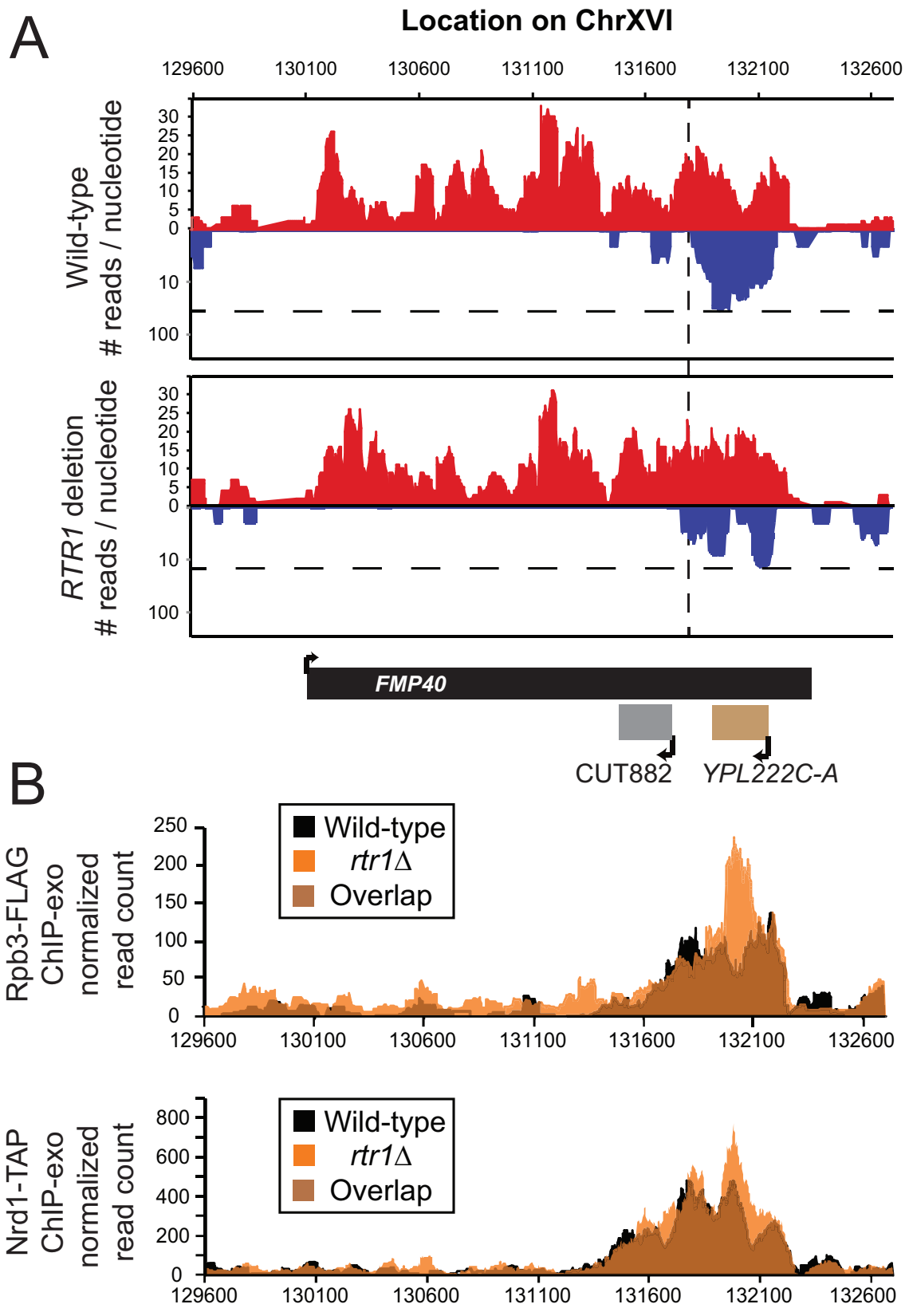
A
Location on ChrXVI
129600
130100
130600
131100
131600
132100
132600
Wild-type
reads / nucleotide
RTR1 deletion
reads / nucleotide
30
25
20
15
10
5
0
10
100
FMP40
CUT882
YPL222C-A
B
Rpb3-FLAG
ChIP-exo
normalized
read count
Wild-type
rtr1Δ
Overlap
250
200
150
100
50
0
Nrd1-TAP
ChIP-exo
normalized
read count
800
600
400
200
0

Figure 32: *RTR1* deletion shortens the NNS-terminated antisense transcript at *FMP40.* (A) Graphical representation of strand-specific RNA-Seq reads mapped to *FMP40* region. Reads mapped to the positive strand are on top in red, while reads mapped to the negative strand are on the bottom in blue. The dotted black line indicates the estimated TSS of the *YPL222C-A* transcript. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Protein-coding mRNAs are in black, dubious ORFs are in tan, CUTs are in gray, and bent arrows indicate direction of the TSS. (B) Rpb3-FLAG (top) and Nrd1-TAP (bottom) occupancy in WT and *rtr1Δ* as determined by ChIP-exo sequencing reads mapped to the same region and aligned to (A). Reads from WT are in black, and those from *rtr1Δ* are in orange.

3.6. Disruption of NNS-pathway termination through *RRP6* deletion prevents AST regulation by *RTR1* deletion

As discussed earlier (Figure 25), the majority of transcripts differentially expressed in *rtr1Δ* cells are down-regulated compared to WT. To compare the effect of *RTR1* in the NNS-dependent termination pathway to the effect of *RRP6*, we looked at the most significantly (p-value < 0.005, FDR < 0.25) down-regulated transcripts in *rtr1Δ* cells. In contrast, many of the antisense transcripts that are differentially expressed in *rtr1Δ* cells are up-regulated in *rrp6Δ* cells. As shown in Figure 33A, the majority of these ASTs have nearly identical expression in the *rtr1Δ rrp6Δ* double deletion as they do in *rrp6Δ* cells, although some transcripts appear similar to the WT strain (Figure 33A). These findings support the hypothesis that Rtr1 and Rrp6 are acting through the same pathway with opposing actions and that deletion of *RRP6* masks the effects of the loss of *RTR1*. This model is illustrated in the specific examples of CUT882, which is transcribed antisense to *FMP40* (Figure 33B), as well as the antisense transcript at *GAP1* (Figure 33C, D). We have previously reported using northern blot analysis that CUT882 is an extended *YPL222C-A* transcript transcribed near the 3'-end of *FMP40* (see Figure 32 for diagram). As mentioned, the CUT882 transcript is shorter in *rtr1Δ* cells and longer in *rrp6Δ* cells as detected by RNA-Seq and Rpb3 ChIP-exo. The annotation for CUT882 is similarly increased in the *rtr1Δ rrp6Δ* double mutant as quantified by RNA-Seq suggesting that *YPL222C-A* is also longer in the double mutant. *GAP1,* encoding a general amino acid permease, also appears to be an antisense-responsive gene [214]. Upon the

loss of *RTR1*, expression of *GAP1* is increased 2-fold and the levels of the AS *GAP1* transcript are decreased by 50% (Figure 33C, D). Conversely when *RRP6* is lost, *GAP1* is decreased by 50% and the AS transcript is increased 1.5-fold. In the *rtr1Δ rrp6Δ* double deletion strain, expression of both transcripts is unchanged from WT cells. These data show that the AS *GAP1* non-coding transcript is differentially regulated by Rtr1 and Rrp6, such that the double mutant effectively rescues the expression of the sense and antisense *GAP1* transcripts.

It has been well documented that *IMD2* is regulated by alternative selection of transcription start sites and NNS-dependent early termination [146, 170, 208, 223]. RNA-Seq data indicates that Rtr1 alters the expression of *IMD2* and its upstream regulatory transcript, a known target of the NNS pathway (Figures 34, 35). We previously showed that Rrp6 is required for proper NNS-dependent termination at specific target gene transcripts [243]. In WT cells, RNA-Seq reads map across *IMD2* and the upstream transcript at similar levels during growth in YPD (Figure 34A). When *RTR1* is knocked out, the expression of both transcripts is significantly decreased (p-value ≤ 5.0×10^{-07}, Figures 34, 35). This is consistent with previous reports that RTR1 deletion mutants are sensitive to the *IMD2* inhibitor mycophenolic acid [185, 244]. The decrease in *IMD2* RNA is confirmed by quantitative PCR (Figure 34B). The ratio of upstream transcript to IMD2 transcript is similar between WT and *rtr1Δ* strains (Figure 34C). When *RRP6* is knocked out, *IMD2* expression is similar to WT levels, but the upstream transcript is significantly increased (p-value ≤ 1.09×10^{-10}, Figure 35A). This is

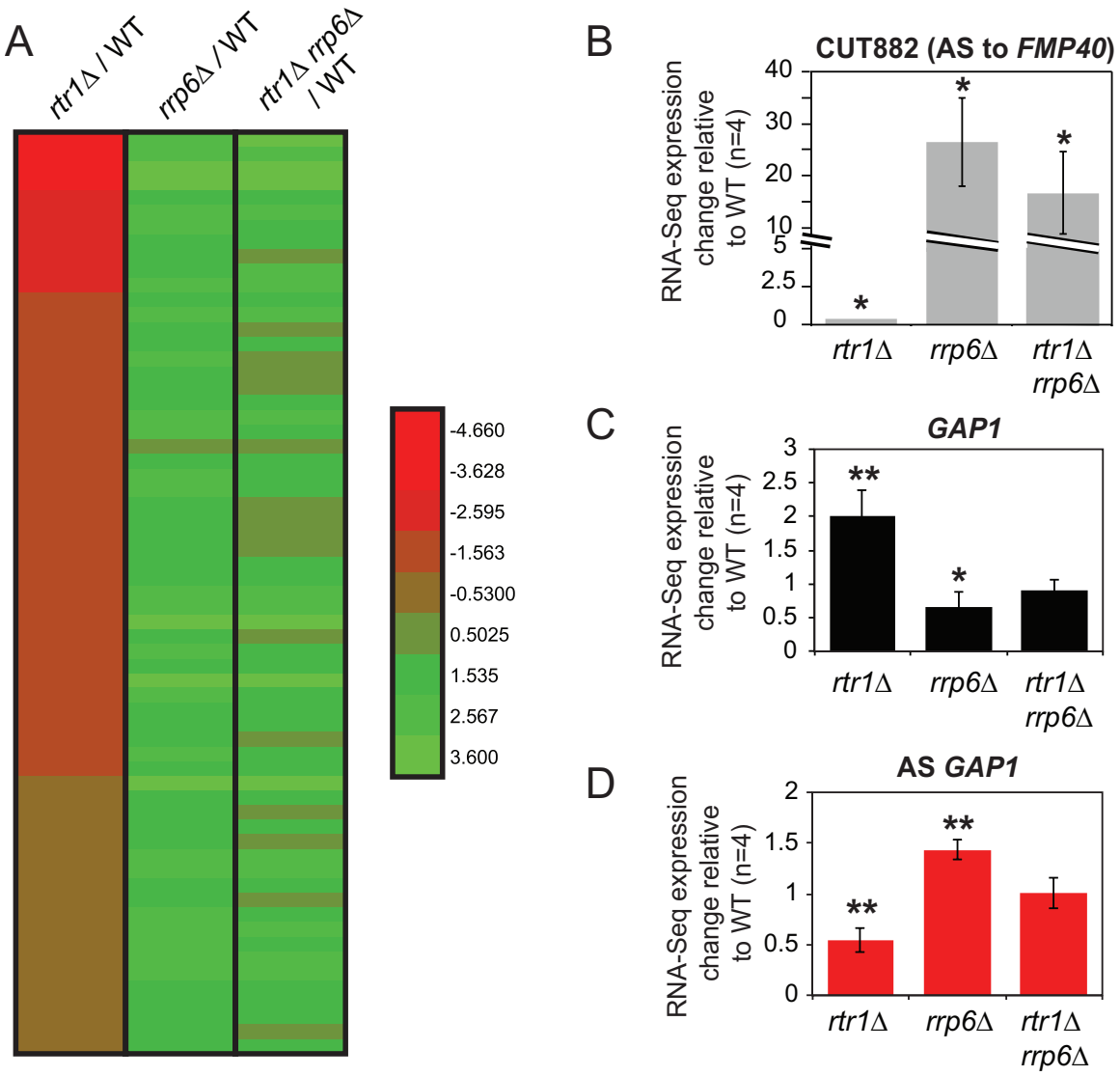

Figure 33: *RTR1/RRP6* double deletion strain phenocopies *rrp6Δ* cells. A) Heat map of transcripts differentially expressed in *rtr1Δ* cells sorted from most decreased in *rtr1Δ* cells compared to WT cells, according to the scale at the right. B-D) Average normalized read counts ± standard deviations for CUT882 (B), *GAP1* (C), and *AS GAP1* (D) in indicated strain versus WT (n=4). Two stars indicate a p-value of <0.01, one star indicates a p-value of <0.05 as determined by an unpaired, two-tailed student's t-test. The colors of the bars in each graph correspond to the color representing the relative annotation.

previously been described. Interestingly, in the *rtr1Δ rrp6Δ* double mutant, the *IMD2* transcript levels are reduced to the low levels seen in the *rtr1Δ* single mutant, but the upstream transcript now appears at levels similar to WT. The amount of upstream transcript is significantly increased in *rtr1Δ rrp6Δ* double mutant strains relative to *rtr1Δ* cells (p-value = 0.003). These data suggest that the upstream transcript is transcribed and then efficiently degraded by the nuclear exosome in the presence of Rrp6. However, the high efficiency of NNS termination in the absence of Rtr1 appears to prevent normal levels of transcription of the protein coding *IMD2* transcript.

To distinguish between effects of RNA degradation versus transcription termination, we also looked at the RNAPII and Nrd1-bound RNAPII by ChIP-exo (Figure 35B, C). In all strains, we see the majority of RNAPII at the upstream transcript. In Rpb3-FLAG strains, RNAPII occupancy at *IMD2* is decreased by approximately 75% in *rtr1Δ* cells. However, in Nrd1-TAP strains, occupancy is only decreased by 50% in *rtr1Δ* cells, indicating that the ratio of Nrd1 to RNAPII is relatively increased in *rtr1Δ* strains as previously shown for other NNS target genes in this study. Overall these data suggest that loss of Rtr1 activity increases Nrd1 binding to RNAPII, most likely due to the increased Ser5-P on the CTD. The increase in Nrd1 recruitment improves the efficiency of termination at the upstream transcript and effectively inhibits transcription initiation at the downstream *IMD2* TSS. When Rrp6 is present, this upstream transcript is degraded, but it is stabilized when *RRP6* is deleted without a change in the

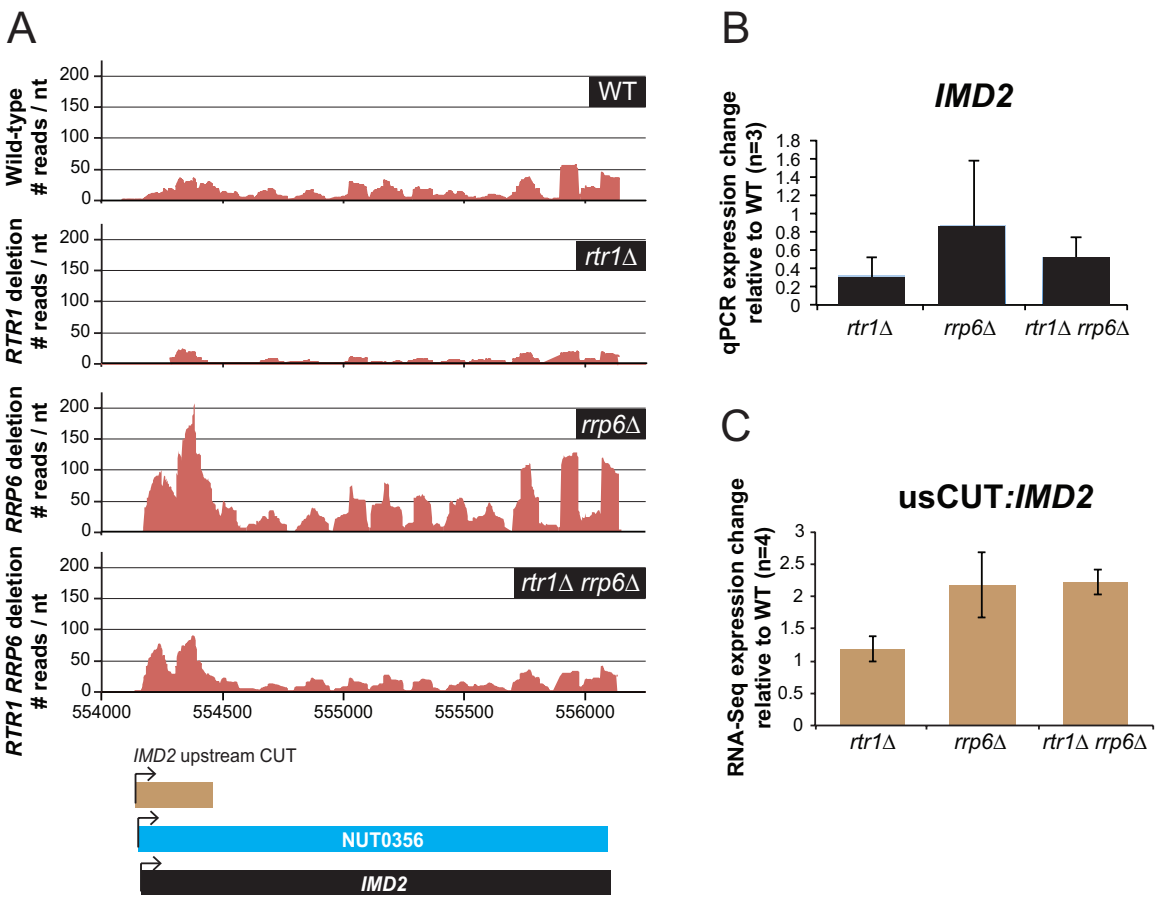

Figure 34: Deletion of *RTR1* disrupts expression of *IMD2* mRNA through *RRP6*. A) Graphical representation of strand-specific RNA-Seq reads mapped to *IMD2* region. Reads mapped to the positive strand are on top in red, while reads mapped to the negative strand are on the bottom in blue. The location and direction of transcription for all analyzed annotations are diagrammed below the graphs to scale. Protein-coding mRNAs are in black, dubious ORFs are in tan, NUTs are in aqua, and bent arrows indicate direction of the TSS. B) Confirmation of expression of *IMD2* by QPCR are indicated as the average fold-change from WT in *rtr1Δ*, *rrp6Δ*, and *rtr1Δ rrp6Δ* strains (n=3). C) The average ratio of RNA-Seq reads mapped to the *IMD2* upstream CUT annotation compared to the number of reads mapped to the *IMD2* gene annotation in *rtr1Δ*, *rrp6Δ*, and *rtr1Δ rrp6Δ* strains relative to WT (n=4).

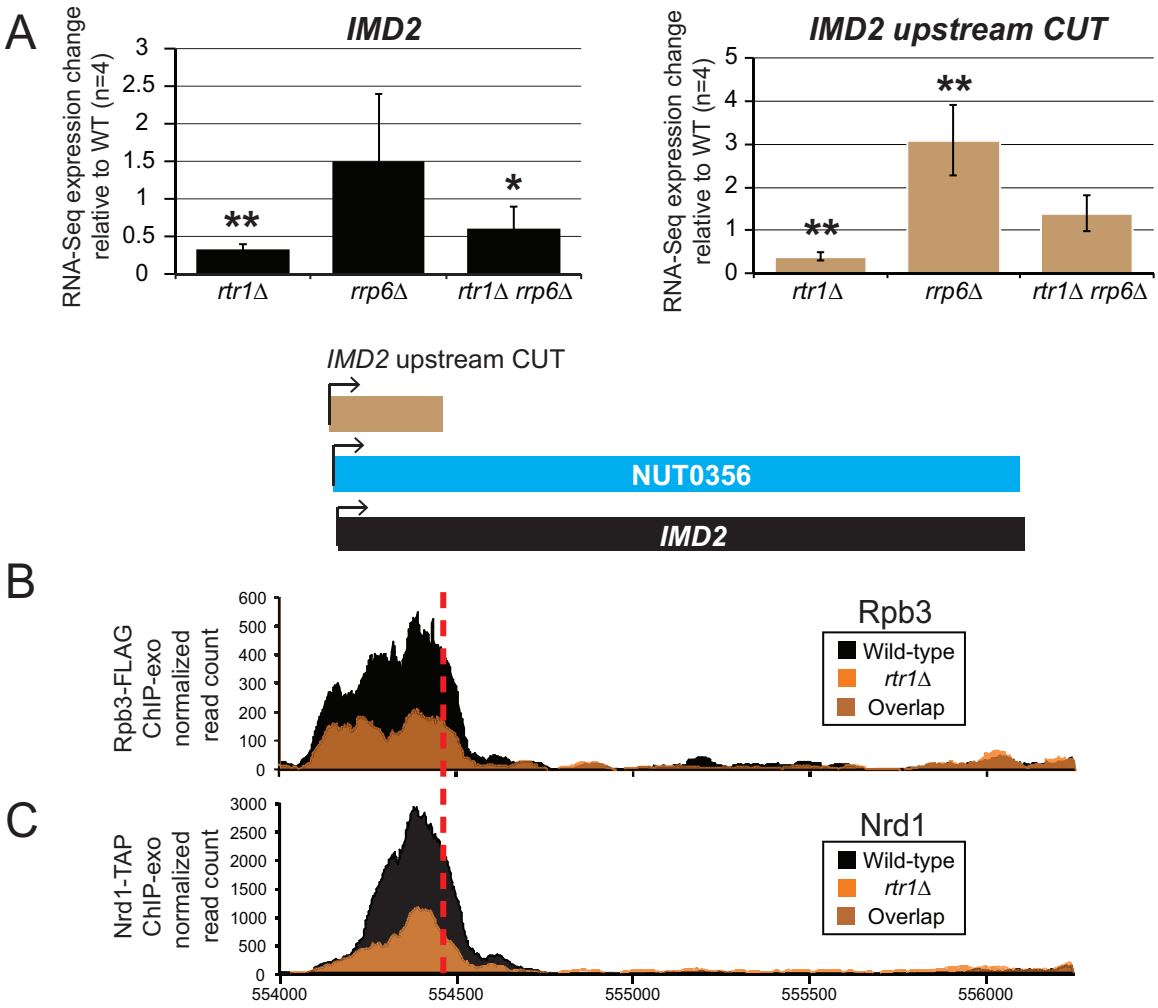

Figure 35: Deletion of *RTR1* disrupts the NNS-dependent transcription termination at the *IMD2* gene through *RRP6*. (A) Average differential expression ± standard deviations for *IMD2* (left) and the *IMD2* upstream CUT in *rtr1Δ*, *rrp6Δ*, and *rtr1Δ rrp6Δ* strains relative to the WT strain (n=4). One star indicates a p-value of <0.05, and two stars indicate a p-value of <0.01 as determined by an unpaired, two-tailed student's t-test. The colors of the bars in each graph correspond to the color representing the related annotation. (B and C) Rpb3-FLAG (B) and Nrd1-TAP (C)) occupancy in WT and *rtr1Δ* as determined by ChIP-exo sequencing reads. The location and direction of transcription for all analyzed annotations are diagrammed above the graphs to scale. Protein-coding mRNAs are in black, NUTs are aqua, the upstream CUT is tan, and bent arrows indicate direction of the TSS. WT reads are in black, and *rtr1Δ* are in orange. The dotted red line indicates the transcription termination site of the *IMD2* upstream CUT.

downstream abundance of RNAPII, suggesting that Rrp6 is not required for termination at the IMD2 upstream CUT.

3.7. Deletion of *RTR1* increases the efficiency of the NNS pathway

Our data indicate that Rtr1 plays a role in the NNS-dependent termination pathway at gene transcripts expressed at low-levels. As mentioned previously, Nrd1 is required for termination of the highly expressed and stabilized snRNAs. Although we do not detect any snRNAs that are altered in length when *RTR1* is deleted, a few snRNAs are expressed at higher levels in *rtr1Δ* cells such as snR65 and snR56 (Figure 36A). Both of these snRNAs are expressed 2-fold higher in *rtr1Δ* cells than in WT cells and decreased by 50% in *rrp6Δ* and *rtr1Δ rrp6Δ* strains. As discussed previously, deletion of *RRP6* may result in improperly processed 3'-ends at these snRNAs, causing destabilization and decreased expression [243]. We hypothesize that deletion of *RTR1* increases the efficiency of NNS-dependent termination at these genes, but Rrp6 is required for proper 3'-end processing and stability in agreement with previous findings. We also see an interesting effect in both our RNA and ChIP data at the RNAPIII transcribed gene snR52. There is a CUT upstream of snR52 previously annotated as CUT567 [72]. Although there is not much difference in the expression of RNA from this CUT in *rrp6Δ* cells, CUT567 and snR52 are both increased 2-fold or more in the *rtr1Δ* strain (Figure 36B). Rpb3-FLAG ChIP-exo data indicates that RNAPII localizes to CUT567 with low, but detectable levels at snR52 in WT, but RNAPII occupancy is increased at snR52 in *rtr1Δ* cells (Figure 36C). This is a rare

example where we see an increase in Rpb3 localization in the *rtr1Δ* strain. Nrd1-TAP also localized to both CUT567 and snR52 in WT and is increased in *rtr1Δ* cells (Figure 36D). One possible explanation is that the loss of Rtr1 leads to increased Nrd1 interaction with RNAPII, and in some instances, this association inhibits release of RNAPII from the DNA. Alternatively, this could indicate that deletion of *RTR1* could alter transcription initiation of RNAPII allowing it to initiate at RNAPIII target genes.

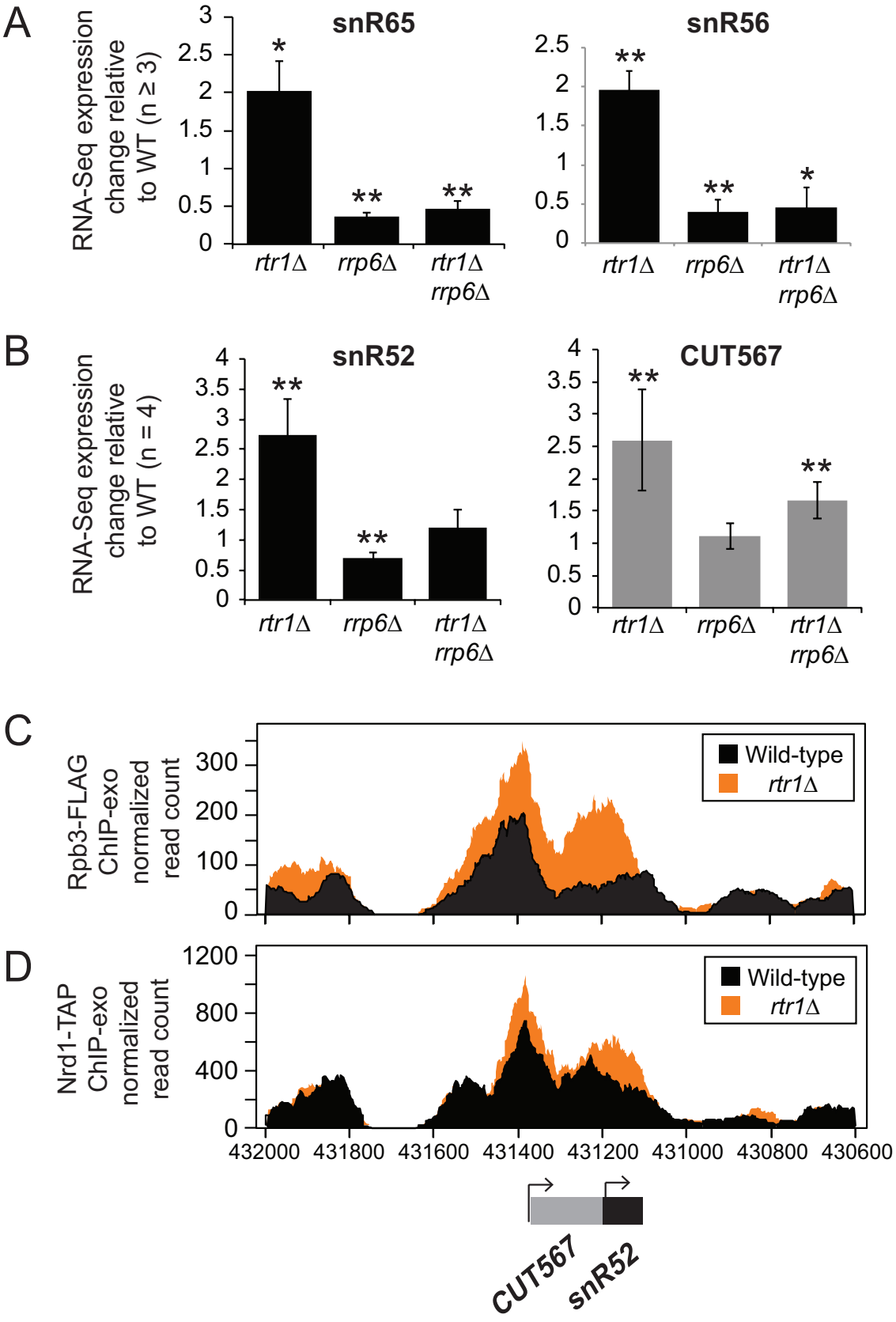
A
RNA-Seq expression change relative to WT (n ≥ 3)
snR65
snR56
rtr1Δ
rrp6Δ
rtr1Δ rrp6Δ
B
RNA-Seq expression change relative to WT (n = 4)
snR52
CUT567
C
Rpb3-FLAG ChIP-exo normalized read count
Wild-type
rtr1Δ
D
Nrd1-TAP ChIP-exo normalized read count
Wild-type
rtr1Δ
432000 431800 431600 431400 431200 431000 430800 430600
CUT567
snR52

Figure 36: Deletion of *RTR1* increases the efficiency of the NNS-dependent termination pathway at some snRNAs. (A and B) Average differential expression ± standard deviations for snR65, snR56, snR52, and CUT567 in *rtr1Δ*, *rrp6Δ*, and *rtr1Δ rrp6Δ* cells relative to WT (n=4). One star indicates a p-value of <0.05, and two stars indicate a p-value of <0.01 as determined by an unpaired, two-tailed student's t-test. The colors of the bars in each graph correspond to the color representing the related annotation. (C and D) Rpb3-FLAG (C) and Nrd1-TAP (D) occupancy in WT and *rtr1Δ* as determined by ChIP-exo sequencing reads. The location and direction of transcription for all analyzed annotations are diagrammed above the graphs to scale. Protein-coding mRNAs are in black, CUTs are gray, and bent arrows indicate direction of the TSS. WT reads are in black, and *rtr1Δ* reads are in orange.

DISCUSSION

1. Homologous mechanisms in mammals

1.1. Structure of the Human Exosome

Although the structure of the human exosome has not been solved, the organization of the core complexes are well conserved with its yeast counterpart (reviewed in [245]). The human RNA exosome is more complex regarding the catalytic subunits and associated cofactors. While there is only one form of the Dis3 nuclease in yeast, there are three in humans, DIS3, DIS3L1, and DIS3L2, each with distinct subcellular localization patterns [94, 95]. By contrast, Rrp6 only has one human ortholog, hRRP6 (also known as PM/Scl100). In the yeast exosome, most of the highly structured RNA substrates are degraded by Dis3 and cannot be processed by Rrp6. In humans, hRRP6 is able to processed more structured RNAs, probably do to the more open conformation assumed by hRRP6 [118]. hRRP6 also differs from its yeast counterpart in that it is most highly concentrated in the nucleolus, but is also present in the nucleoplasm and cytoplasm [69]. Human homologs of subunits of the TRAMP complex have also been identified, including Mtr4 and the Trf and Air proteins [155, 246, 247]. Interestingly, a second distinct complex has been identified to have TRAMP-like functionality but is restricted to the nucleolus. This complex, called the Nuclear Exosome Targeting Complex (NEXT), is made up of hMTR4 and two other proteins, ZCCHC8 and RBM7 [247]. NEXT has been found to play an important role in targeting pre-snRNA transcripts to the exosome for processing [248].

1.2. Human Exosome Interacts with Capping Complex

A number of studies report links between the nuclear exosome and the Cap Binding Complex (CBC) (Figure 5). The 5' mRNA cap is involved in coordinating a number of mRNA processing events such as splicing [249], 3' end formation [250], and RNA turnover [251], but specifically how cap formation fulfills these roles is unclear. Regarding RNA turnover by the nuclear exosome, it appears that proteins associated with the exosome may form a physical link between the exosome and CBC. In yeast, it has been shown that Rrp6 directly interacts with Nrd1, which in turn interacts with CBC proteins [70]. In humans, the exosome cofactor the Nuclear Exosome Targeting complex (NEXT) has been shown to interact with the CBC [252, 253]. Furthermore, it has been shown in yeast that aberrant mRNAs that cannot be exported from the nucleus are subjected to degradation by a process requiring both Cbc1, a nuclear cap-binding protein, and Rrp6 [254, 255]. The fact that the nuclear exosome and Nrd1 may both interact with the CBC [70] is interesting. It is possible that the CBC plays an important, unknown role in NNS-dependent termination at specific genes, and this role may require Rrp6 leading to the effects detected in our studies.

2. Implications of Transcription Events in Human Health and Disease

2.1. Transcription Regulation by Termination

Altered termination of RNAPII transcription termination can affect gene regulation and protein expression in at least four ways, described at length below: 1)

alternative polyadenylation site usage, 3) defective 3’ end processing, 3) early transcription termination of protein-coding mRNAs,and 4) regulation of cryptic and antisense transcription.

Alternative Polyadenylation Site Usage

It has been reported that upwards of 70% of human pre-mRNAs and lncRNAs have more than one polyadenylation signal sequence [256]. Orchestration of specific cleavage and polyadenylation factors at *cis*-elements play major roles in the subsequent choice for the site of polyadenylation of the mRNA [257]. Alternative polyadenylation (APA) changes the length of the 3’-UTR and can quantitatively effect the expression of the gene by altering the availability of protein-binding sites and/or the composition of regulatory sequences [258, 259]. To illustrate this idea, it was reported that alternate 3’-ends of RNA may regulate localization of the membrane proteins CD44, CD47, ITGA1, and TNFRSF13C [260]. In the case of *CD47*, the extended 3’ UTR serves as a co-translational scaffold for RNA-binding proteins that facilitate recruitment of proteins required for translocation to the plasma membrane. CD47 proteins translated from an mRNA with a short 3’ UTR localized instead to the endoplasmic reticulum.

APA can also play a role in cell differentiation, and APA patterns may be tissue specific [261]. During development of the central nervous system, the expression of two isoforms of GFAP (glial fibrillary acidic protein) is time-dependent and regulated by APA. The exact mechanism of regulation is not

known, but splicing factors may be involved [262]. In another example, one aspect of maturation of B cells may be regulated by APA in a related mechanism. Young B cells express immunoglobulins bound to the cell surface, while mature cells secrete them. This isoform shift is regulated by splicing-dependent APA in which the splicing and 3' end processing machinery compete [263-265]. APA may regulate developmental processes on a more global scale in a tissue-specific fashion during spermatogenesis and brain development [266-270]. Generally, more distal polyadenylation sites are used increasingly as differentiation progresses, resulting in the inclusion of more miRNA and protein binding sites [271-273]. Additionally, proliferating cells, including gastric cancer cells [274], tend to use earlier polyadenylation sites and, therefore, have shorter 3'-UTRs [275, 276].

Processing of 3' ends of RNA

Proper 3'-end processing is important for regulation of gene expression and function. As highlighted above, it has become clear that sequences in 3'-UTRs and their recognition by polyA binding proteins play an important role in the majority of cellular processes studied [277]. Improper mRNA 3'-end processing has been associated with a number of diseases (reviewed in [278]. Proper protein expression may be dependent on sequences in the 3'-UTR of the mRNAs, such as sequences containing recognition sites of microRNAs [279]. A few human mRNAs have non-canonical 3'-end processing signals called upstream sequence elements (USE sites) [280-283]. Regulated utilization of

these sites enhance 3'-end processing at otherwise weak signals [284]. One such protein, the serine protease thrombin, is an important regulator of blood coagulation and clotting. Its expression is regulated in response to inflammation and stress by recognition of specific sequences in the 3'-UTR of mRNAs [285, 286]. Improper utilization of the USE site leads to abnormal coagulation and blood clots [285, 287]. Additionally, a mutation in the polyadenylation signal of the tumor suppressor gene TP53 present in 0.5-2% of European populations has been linked to a predisposition to a number of cancers [288, 289]. Mutations in the FOXP3 (forkhead box P3) polyadenylation signal can cause IPEX syndrome, a complex syndrome characterized by immune dysfunction, polyendocrinopathy and enteropathy [290]. FOXP3 is a transcription factor in T cells, and the 3'-UTR resulting from the mutation is very long making the protein unstable and decreasing its expression.

Early Termination of Protein Coding RNA and Regulation of Cryptic Transcription

The NNS-dependent termination pathway is an important regulator of gene transcription. In examples such as *NRD1*, *HRP1*, and *URA2*, transcription can be terminated prematurely by the Nrd1 termination pathway [43, 145, 152, 170, 223]. Only transcripts that escape this early termination can produce functional mRNAs, and alterations that increase the effectiveness of the NNS pathway lead to a decrease in mRNA expression. At some genes, noncoding RNAs terminated by Nrd1 overlap with the promoter of, or are antisense to, an adjacent protein-coding gene [153, 170, 214]. For instance, *IMD2* gene expression is regulated by

upstream transcription start sites used when guanine levels are high. These start sites encode short, unstable transcripts that are terminated by the NNS pathway [146, 170, 208]. In other instances, protein-coding genes are regulated by NNS-dependent termination of an antisense non-coding RNA, such as *FMP40* and *PHO84* [41, 50, 217, 231]. At these noncoding genes, antisense transcription is initiated at the 3'-end of the protein-coding gene and terminated by the NNS pathway after a short distance. If NNS-dependent termination does not occur, transcription of the antisense strand will continue through the entire gene and over the transcription start site of the protein-coding gene, inhibiting proper initiation at the mRNA [217, 243].

2.2. Transcription Addiction in Tumor Cells

Many human cancers depend on deregulated transcription of oncogenes and tumor suppressors. In some cases, cancer cells may be "addicted" to the overexpression of a single oncogene [291]. For instance, it has been well documented that the Bcr-Abl fusion protein drives proliferation in chronic myelogenous leukemia (CML). Flavopiridol, a drug that inhibits transcription by targeting the RNAPII CTD kinase P-TEFb (CTK9) [292] may be used in conjunction with the Bcr-Abl inhibitors to decrease imatinib resistance in CML [293]. In other cases, the cancer cells may contain a mutation in an oncogene that activates a number of downstream targets. Over- expression of the *MYC* gene family aids in aggressive proliferation and poor patient outcomes [294, 295] by causing a cascade of global transcription amplification and up-regulation of

genes involved in multiple processes [296-299]. It has recently been reported that the proliferative effect of up-regulation of the *MYCN* oncogene can be suppressed in neuroblastoma cells by inhibition of the cyclin-dependent kinase and general transcription factor CDK7 with limited toxicity to normal cells [300]. CDK7 is the human homolog of Kin28, the catalytic subunit of TFIIH that adds the Ser5-P mark later removed by Rtr1, and is the major protein kinase that modifies Ser5 and Ser7 in the RNAPII CTD in human cells [301]. These findings support a working model in which some cancer cells become “addicted” to the amplified transcription of proliferative genes. A better understanding of general transcription factors and mechanisms is needed to determine if it is possible to target transcription addiction in cancer cells with limited effects in normal cells.

2.3. Senataxin, the Human Homolog of Sen1

The RNA:DNA helicase Sen1 and its homolog in humans Senataxin contain a conserved helicase domain [302, 303]. Mutations in Senataxin are associated with the juvenile neuromuscular disorders ataxia oculomotor apraxia type 2 (AOA2) and amyotrophic lateral sclerosis type 4 (ALS4) [304-306]. The function of Senataxin mutations in these disorders is unknown, but many are located within the helicase domain [302]. This suggests that the helicase function may play a role in neurogenerative diseases.

Studies of Sen1 have revealed that its helicase function is important for protecting against genomic instability and damage [307-309]. As RNAPII transcribes a gene, the nascent RNA can bind to the complementary template

DNA strand forming an RNA:DNA hybrid with the single stranded non-template DNA strand left exposed in a structure called an R-loop. Without removal of the R-loop by Sen1, the single stranded DNA is more susceptible to mutation and recombination [310-315]. The R-loop structure can lead to RNAPII stalling which inhibits transcription elongation and can lead to transcription fork collapse when a replicating DNA Polymerase encounters a stalled RNA Polymerase [316-320]. Furthermore, Sen1 may play a role in regulating the expression of the DNA repair gene RNR1 [309] and an inactivating mutation in the Sen1 helicase domain has been shown to alter RNAPII localization genome-wide [53]. Sen1 has also been shown to play a direct role in DNA damage response [86, 307, 308, 321, 322], and Senataxin forms stress-induced foci that may localize at chromosome fragile sites or repetitive sequences, either of which may be susceptible to R-loop formation [321, 323, 324]. It is also possible that brain-specific Senataxin protein-protein interactions are disrupted by the disease-associated mutations. Richard and Manley have reported that Senataxin is normally sumoylated and interacts with the core exosome subunit Rrp45 via the sumoylation [325]. AOA2-linked mutations inhibit Senataxin sumoylation, but ALS4-linked mutations do not [86, 325].

2.4. Rrp6 and the Nuclear RNA Exosome

Autoimmune Disease

Antibodies against many of the subunits of the human exosome have been characterized in autoimmune disorders affecting the connective tissue,

specifically: PM (polymyositis), Scl (scleroderma) and PM/Scl overlap syndrome [83, 326-328]. Patients' sera were found to contain antibodies against hRrp6 (called PM/Scl-100 at the time), and some patients also had antibodies against hRrp45, one of the RNase PH-like core proteins (called PM/Scl75) [98-100]. The role of the exosome in these disorders remains unclear.

Neurological Disorders

Two of the core subunits of the human exosome, EXOSC3 (homolog to the yeast cap protein Rrp40) and EXOSC8 (homolog to the yeast barrel protein Rrp43), have been implicated in pontocerebellar hypoplasia type 1 (PCH1) and spinal motor neuron degeneration in several families [84, 102]. PCH is a group of autosomal recessive disorders characterized by progressive deteriorations of various parts of the brain. Patients with PCH1 also have diffuse muscle wasting. Mutations in EXOSC8 have also been linked to cerebellar hypoplasia and spinal muscular atrophy [85]. The reporting study determined that down-regulation of EXOSC8 in human cells and zebrafish resulted in down-regulation of myelin encoding genes and an imbalance of myelin protein. These and other studies involving mRNA processing factors indicate that proper regulation of mRNA processing is critical in neuromuscular development and maintenance [86, 321, 329-331]. Furthermore, non-coding RNAs appear to play important roles in neurological development, and disregulation of non-coding RNAs has been found to lead to neurodegeneration [331-333]. Proper processing, quality control, and

degradation by Rrp6 and the nuclear exosome may serve important functions keeping non-coding RNAs in check in neuromuscular tissue.

Cancer

It is not clear if Rrp6, the nuclear exosome, and their cofactors play a direct role in cancer formation and progression, but they have been linked to DNA repair and chromatin restructuring mechanisms that are important for maintaining genomic integrity. The TRAMP complex may play a role in chromatin remodeling and DNA repair, important for maintaining properly regulated gene expression and inhibiting cancer growth [227]. TRAMP may be functionally linked to RNAi machinery in S. pombe and aid in post-transcriptional gene silencing and RNA induced transcriptional silencing which result in heterochromatin formation over the target gene [334]. One early study suggested that the core exosome subunit Rrp46 may be up-regulated in 10–33% of patients with lung cancer, melanoma, and prostate cancer [101]. Furthermore, in a genome-wide drug-target screen, Rrp6 was found to be a potential target of the cell growth inhibitor 5-fluorouracil used in chemotherapy for the treatment of solid tumors. The role of Rrp6 in rRNA processing is proposed as a pathway by which 5-fluorouracil inhibits growth [335]. Rrp6 may play a role in repair of DNA double strand breaks through the homologous recombination pathway in *Drosophila* and humans [336]. Marin-Vicente *et al.* recently reported that Rrp6 is recruited to double strand breaks in *Drosophila* S2 cells and human HeLa cells and co-immunoprecipates with Rad51, a protein involved in recombination. Both deletion

of RRP6 and overexpression of a catalytically inactive Rrp6 mutant disrupt recruitment of RAD51 to the damage sites but does not decrease Rad51 protein levels. Furthermore, depletion of Rrp6 makes both cell lines sensitive to radiation, indicating Rrp6 may be involved in DNA damage repair [336].

3. Transcription termination effects gene regulation

3.1. Exosome aids in backtracking RNAPII

Although most studies have implied that Rrp6 and the exosome targets RNA substrates after they are released from RNAPII, the core exosome can associate with RNAPII on the DNA. It is known that the exosome associates with RNAPII on target genes in *Drosophilia* [337]. As mentioned above, both catalytic subunits of the nuclear exosome are 3' – 5' exonucleases. As such, the nascent RNA being actively elongated by RNAPII would not be accessible to the exosome. However, recent work has shown that the core exosome can aid in terminating backtracked RNAPII at protein coding genes in *S. pombe* [338]. RNAPII has been shown to backtrack, or reverse directions while still on the transcribed DNA sequence to aid in RNAPII release from the DNA when stalled [339]. By backtracking, RNAPII could expose the 3' end of the RNA, allowing access and degradation by the exosome (model shown in Figure 37) [340]. It was determined that termination of backtracked RNAPII was carried out by Dis3, with no detectable effect from an *RRP6* knockout. However, this study focused on protein coding mRNAs. Rrp6 may play a greater role in rescuing backtracked

RNAPII transcribing noncoding RNAs, as Rrp6 has been shown to play a greater role in termination of cryptic transcripts originating from within genes [172].

3.2. Regulation of PolyA Tail Length

The length of the polyA tail added to pre-mRNAs after transcription affects the stability and translation efficiency of the mRNA (Reviewed in [341]). The normal polyA tail length (about 70 – 80 nt in *S. cerevisiae*) is carefully regulated by interactions between PolyA Binding Proteins (PABPs) and polyA polymerases. Two PABPs have been identified in yeast: Pab1 which is primarily in the cytoplasm but can be shuttled into the nucleus [342-344], and Nab2 which is primarily in the nucleus but is exported with the mRNP into the cytoplasm where it is thought to be exchanged for Pab1 [345-348]. Expression of Nab2 is subject to post-transcriptional auto-regulation in a model in which Nab2 binding to the 3' UTR of a *NAB2* mRNA results in recruitment of Rrp6 and subsequent mRNA degradation [349]. It has been reported that Rrp6 can physically interact with Nab2, displacing Nab2 from polyA tails, leaving the unprotected 3'-ends available for subsequent degradation [77].

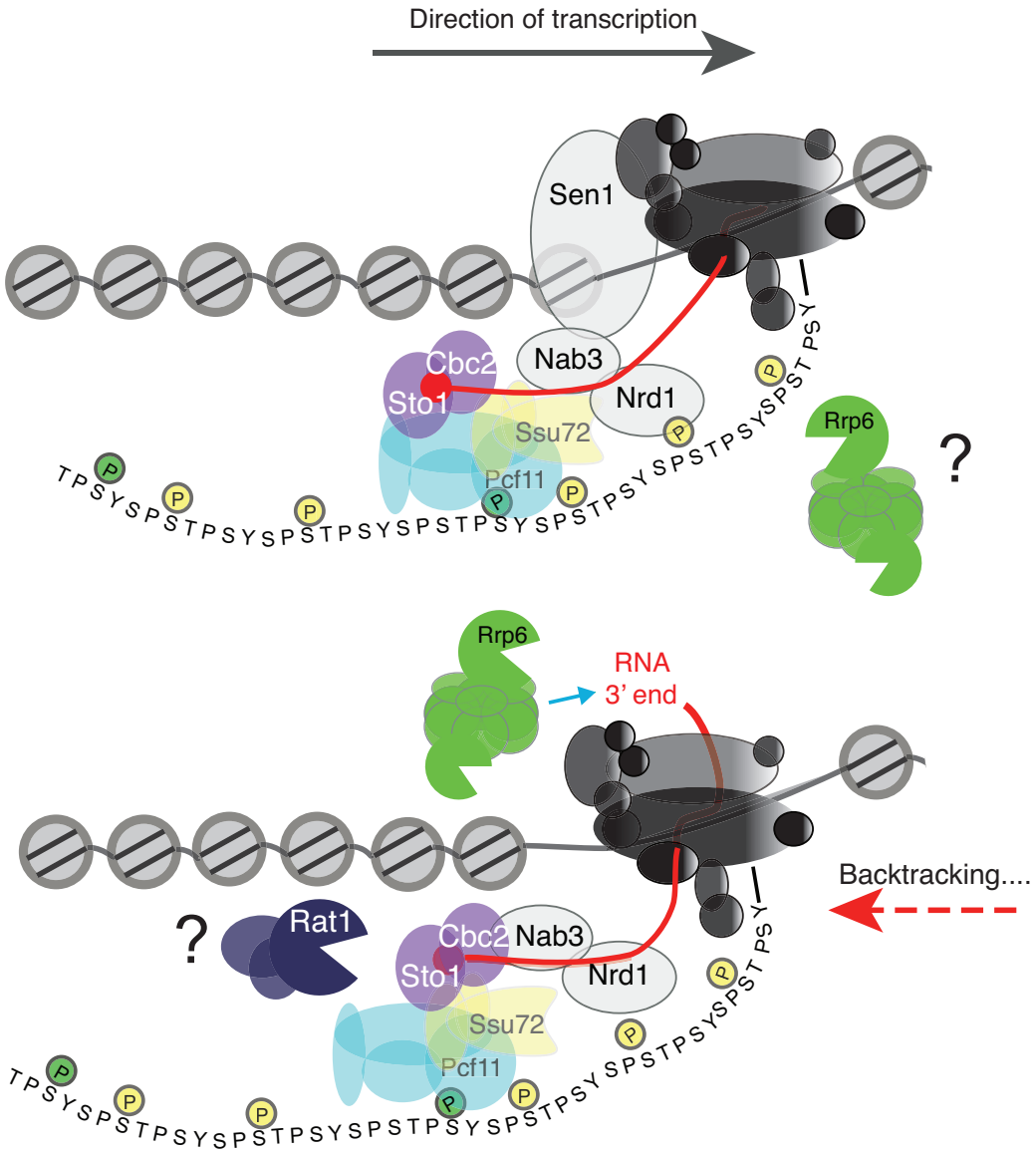

Figure 37: Model for how the nuclear exosome may facilitate rescue of backtracked RNAPII. Dis3 and the core exosome have been found to facilitate termination of backtracked RNA, but Rrp6 is suggested not to play a significant role in terminating backtracked mRNA transcripts in *S. pombe*. Rrp6 is suggested to function in termination of cryptic antisense transcripts originating within genes [172] and certain snRNAs [243], so it is possible that Rrp6 assists backtracking RNAPII for these gene transcripts by a model similar to that suggested for Dis3. When RNAPII is transcribing, the 3' end of the RNA is in the active site, protected by RNAPII, and inaccessible to Rrp6. When RNAPII backtracks in response to an obstacle in the DNA or improperly incorporated nucleotide, the 3' end of the RNA emerges through a 'backtrack site' in RNAPII and can be accessed by Rrp6 [340].

4. Additional Roles of the RNA Exosome in Nuclear RNA processing and Surveillance

4.1. RNA Turnover

With transcription occurring across so much of the genome, the cell must recover ribonucleotides tied up in these transcripts. Although many transcripts must be stabilized to perform their functions, such as mRNAs, rRNAs, and tRNAs, errors in transcription and processing may result in aberrant transcripts that need to be recycled. Furthermore, certain transcripts appear to be transcribed with no known purpose, other than possibly to regulate the transcription of nearby regions *in cis*. The nuclear exosome appears to recognize these unstable transcripts and targets them for degradation, recycling the ribonucleotides contained within [141-143, 156, 157, 159, 350, 351].

4.2. pre-mRNA

Rrp6 and the exosome are involved in quality control and turnover of mRNAs at several steps including splicing and 3'-end processing. The splicing machinery recognizes intron-containing mRNAs co-transcriptionally. Once these transcripts are bound by the spliceosome, they are restricted to the nucleus until they are released [352, 353] (Figure 6C, left panel). This finding suggests that any pre-mRNAs bound to the spliceosome that cannot be properly spliced must be degraded in the nucleus. Bousquet-Antonelli *et al.* discovered that these pre-mRNAs are degraded primarily by Rrp6 and the exosome, and to a lesser extent by the 5' – 3' exonuclease Rat1 [354]. Interestingly, they also found that inhibition

of the degradation pathways could increase the levels of certain spliced mRNAs, indicating that pre-mRNAs may compete for splicing and degrading machineries, with some transcripts being favored. This competition and resulting pre-mRNA preference appears to be regulated by metabolic events such as available carbon source and therefore may be an important step in controlling gene expression levels [354].

The exosome appears to play a role in a type of checkpoint for proper 3' end formation (Figure 6C, right panel). Cells with defective Pap1, the poly(A) polymerase that adds the poly(A) tail to stabilize mRNAs, accumulate nascent unpolyadenylated pre-mRNA transcripts at the site of transcription [355]. Of interest, defects in nuclear export machinery also leads to accumulation of nascent RNAs in proximity to the site of gene transcription. In this case, nascent RNAs become hyperadenylated with poly(A) tails approximately 30 nt longer than expected [356, 357]. Surprisingly, Rrp6 was found to be important in the retention of both the hypo- and hyperadenylated transcripts at the site of transcription resulting from these polyadenylation and export defects [355]. Furthermore, it has been shown that mRNAs with 3' extensions due to improper termination can be processed by Rrp6 to a functional length [358]. Deletion of *RRP6* is sufficient to see accumulation of aberrant transcripts caused by a variety of transcription elongation and termination defects [76, 221]. Because of this accumulation of defective transcripts, *RRP6* knockouts have become an important tool in the study of RNAPII transcription elongation and termination.

4.3. Exosome activities in the cytoplasm

As mentioned above, the RNA exosome is also present in the cytoplasm, minus Rrp6. In the cytoplasm, Dis3 is the only exonuclease in the complex. The cytoplasmic exosome interacts with the SKI complex of helicases that targets RNA substrates to the exosome in a manor similar to the TRAMP complex in the nucleus [359]. The SKI complex is only in the cytoplasm in yeast and is required for mRNA turnover and mRNA surveillance by the exosome [360-363]. The cytoplasmic exosome and SKI complex are required for proper mRNA surveillance via the non-stop decay (NSD) pathway [362, 364, 365]. NSD targets mRNAs that do not contain a translation stop codon [364, 366]. Generally, the nuclear exosome specializes in RNA 3'-end processing, rapid degradation of cryptic transcripts, and surveillance of improper RNA processing, while the cytoplasmic exosome specializes in surveillance of mRNAs and RNA turnover. It has been reported that a yeast cell can transcribe upwards of 2,000 rRNA molecules a minute [367] in addition to other noncoding RNAs and mRNAs. Therefore, dual localization of exonucleases such as the exosome is important for efficient surveillance and degradation in the nucleus to conserve energy required for the nuclear export and potential translation of RNAs in the cytoplasm.

4.4. Exosome activity is important for cellular differentiation

Regulation of transcription termination and degradation of transcription products have been shown repeatedly to be required for proper expression at

specific genes. Importantly, studies are beginning to emerge that reveal the importance of RNA exosome processing at the cellular level. In yeast, the nuclear exosome is involved in facilitating meiosis. Budding yeast have been shown to degrade Rrp6 as they enter meiosis, leading to the accumulation of noncoding transcripts termed meiotic unstable transcripts (MUTs) [81]. Degradation of Rrp6 occurs in the first hours of meiosis and corresponds to the timing of RNA replication and induced double strand breaks [81]. However, cells lacking Rrp6 rarely entire meiosis. In what may be a related process, fission yeast express meiotic genes that are degraded by the exosome to inhibit meiosis [368]. Furthermore, it has recently been found that Rrp6 and the nuclear exosome facilitate proper meiotic recombination [82]. The mechanisms regulating the expression of Rrp6 during early meiosis is not clear.

The exosome is also involved in regulating blood cell maturation [369]. Expression of EXOSC8, the human homolog of the exosome core protein Rrp43, is repressed by erythropoiesis regulator proteins GATA-1 and Foxo3. Loss of EXOSC8 or other exosome components can induce erythroid cell maturation. Alternatively, increased exosome activity can repress the expression of genes activated by GATA-1 and Foxo3 [369]. The mechanisms by which the exosome represses expression of GATA-1/Foxo3-regulated genes has not yet been addressed, and it is unclear what specific roles Rrp6 may play. It is possible that the RNA exosome plays an important negative regulatory role in many maturation and developmental processes.

5. Rrp6 is required for proper NNS-dependent termination at specific transcripts

Through comparison of RNA-Seq and RNAPII ChIP-exo datasets, these data show that NNS-terminator read-through occurs at a significant number of NNS-target genes in the absence of the 3-5' exonuclease Rrp6. These findings support the hypothesis that the NNS pathway requires Rrp6 function for both 3'-end processing and regulation of NNS-dependent termination of specific transcripts. Overall, there is a striking similarity between the annotations for Nrd1 unterminated transcripts (NUTs) at many locations and the transcripts observed following deletion of the nuclear specific 3'-5' exonuclease subunit Rrp6 genome-wide. The tight coupling of the NNS pathway with exosome function has been previously characterized [4, 47, 56, 58, 70], however our findings are the first to demonstrate that Rrp6 function is required to regulate termination through the NNS pathway at specific gene transcripts and to show changes in RNAPII localization that are caused by *rrp6Δ*. This finding is in contrast to recent studies that determined that NUT transcripts were significantly longer than Rrp6-dependent transcripts (previously named as CUTs) [214]. However, there are specific instances in which NUTs have been shown to be longer than Rrp6-dependent transcripts including our findings at *FMP40 / YPL22C-A* [214].

Previous studies have found other extended CUT transcripts (eCUTs; specifically CUT060 and CUT095) to be significantly longer in Nrd1 mutants than in *rrp6Δ* cells as well [221]. This was also true for snR13 in agreement with previous results obtained by northern blotting and other approaches [20, 43, 193,

218, 220]. The mechanisms that underlie the specific requirement for Rrp6 at some transcripts for Nrd1 termination are unknown, but we speculate that Rrp6 regulates NNS-termination efficiency through its interaction with Nrd1 [45, 70]. Our analysis also suggests that the increased dynamic range and resolution provided by RNA-Seq and ChIP-exo gives a distinct advantage in precise mapping of the transcripts that accumulate following deletion of *RRP6*. We propose that it is this difference that led to the findings that NUTs were significantly longer than CUTs, since all previous CUT studies were performed using tiling array based platforms, whereas the NUT studies were performed by deep sequencing [72, 73, 214]. Additionally, comparison of the changes in RNA transcript signals to actual changes in RNAPII occupancy allowed us to distinguish between effects of RNAPII termination events versus exonuclease-dependent RNA processing, which was of particular importance for this study.

In this study, we performed multiple biological replicates (n=4) to provide highly accurate differential expression analysis for all classes of annotated transcripts. In addition, we included differential expression analysis of 1215 antisense transcripts that were significantly changed in *rrp6Δ* cells*,* the majority of which (76%) were up-regulated. These data will serve as a valuable resource to determine the diverse roles of the NNS pathway and Rrp6 in transcriptome-wide gene expression regulation. Using RNA-Seq, we have determined that loss of Rrp6 leads to decreased expression of the majority of sn/snoRNA transcripts. Additionally, we have found that the mRNA transcripts from ribosomal protein coding genes show a strong dependence on Rrp6 for control of their steady state

transcript levels. Previous ChIP studies on Nrd1 have shown that although Nrd1 localizes to many ribosomal protein-coding genes, it does not directly bind to the mRNA transcripts from those genes [50]. This class of mRNA transcripts may be regulated by Nrd1 and Rrp6 due to their short length and hence high levels of Ser5-phosphorylated RNAPII at their 3'-ends, which could facilitate interaction with Nrd1 and regulation of termination [3, 4, 6, 45]. By comparing our data to recent 4tU-Seq datasets involving Nrd1-depletion from the nucleus, we found that multiple ribosomal subunit mRNAs display transcript extension following loss of Nrd1 activity. These data strongly suggest that ribosomal protein coding mRNAs may require Rrp6 and the NNS pathway for proper 3' processing and transcript stability.

High-resolution RNAPII occupancy maps generated by ChIP-exo provide unique insights into the mechanisms of NNS-dependent termination. At *NRD1, HRP1*, and *YPL222C-A*, deletion of *RRP6* lead to a 3' shift in RNAPII localization indicating that NNS-dependent termination was delayed or less efficient (Figures 21, 22, 23). This phenomenon was also observed to a lesser extent at *SRG1-SER3* (Figure 23). Interestingly, we observed distinct accumulation of RNAPII at these regions that occurred just 3' to Nrd1 and/or Nab3 binding sites that have previously been mapped by RNA-protein crosslinking approaches (Figure 21) [50, 60]. These findings suggest that termination through the NNS-pathway may be an inefficient process requiring clusters of Nrd1/Nab3 binding sites to for facilitate higher order recruitment of multiple Nrd1-Nab3 heterodimers to carry out transcription termination, an idea that is also supported by other studies [40,

222]. This hypothesis is further supported by a report that showed that NNS-termination occurs in a ‘termination zone’ rather than a specific termination site as proposed for polyA-dependent termination pathways [51]. Through this mechanism, gene expression attenuation at genes such as *NRD1* and *SRG1-SER3* would remain leaky, allowing for transcription of a small percentage of full-length transcripts even in the presence of high Nrd1-Nab3 protein levels, which is what we observe in WT cells at these genes (Figures 21, 22, 23). The requirement for the nuclear exosome and Rrp6 for efficient NNS-regulated attenuation adds a layer of complexity to this process and another path for regulation for these tightly controlled genes.

It is unclear at this time what role Rrp6 has in regulating RNA expression at this specific class of protein-coding genes. Rrp6 may play a greater role in rescuing backtracked RNAPII at noncoding RNAs, as Rrp6 has been shown to play a role in termination of specific snRNAs [43] and cryptic transcripts originating from within gene bodies [147]. The evidence presented here and by others points toward a mechanism in which Rrp6 is directly involved in NNS-dependent termination, but it is possible that there is another explanation for these findings. For instance, there could be some as-of-yet unknown lncRNA that is normally degraded by Rrp6 but, when stabilized, acts to interfere with the activity of one or more proteins required for proper NNS-dependent termination (Figure 38, bottom panel). This mechanism could be similar to that of NEAT1, a long non-coding RNA that is thought to sequester RNA-binding proteins in subnuclear bodies called paraspeckles (reviewed in [370]). In this theoretical

indirect model for Rrp6 function in RNAPII termination, the lncRNA could bind to the RNA Recognition Motifs (RRM domains) of Nrd1 or Nab3, or an unidentified protein. This would not result in decreased transcription, but would nonetheless inhibit termination activity through the NNS pathway.

6. Insights from transcriptome-wide sequencing

6.1. Annotation boundaries vs. transcription boundaries

Transcriptome-wide RNA analysis such as RNA-Seq has recently revealed that most of the eukaryotic genome is transcribed. Previous attempts to annotate the yeast genome have relied on determining the sequences of stable, easily detectable proteins and therefore only report coding regions in the genome. Even updated sequencing-based approaches to annotating the entire yeast genome have difficulty determining the exact TSS and TTS of every transcript [219]. Our findings support the notion that transcriptional boundaries may not be precise for every gene. As mentioned above, many genes like *IMD2* are regulated by alternative TSSs [208]. Additionally, regulation of transcription elongation and termination may alter the location of the TTS, such as we have shown here. It is important to keep in mind that many groups report new annotations that have been estimated by methods that do not accurately and specifically measure the 5'- and 3'- ends of RNAs [72, 214]. Genomic annotations are necessary and important tools that provide points of reference and methods of expression analysis, but they are difficult if not impossible to perfect.

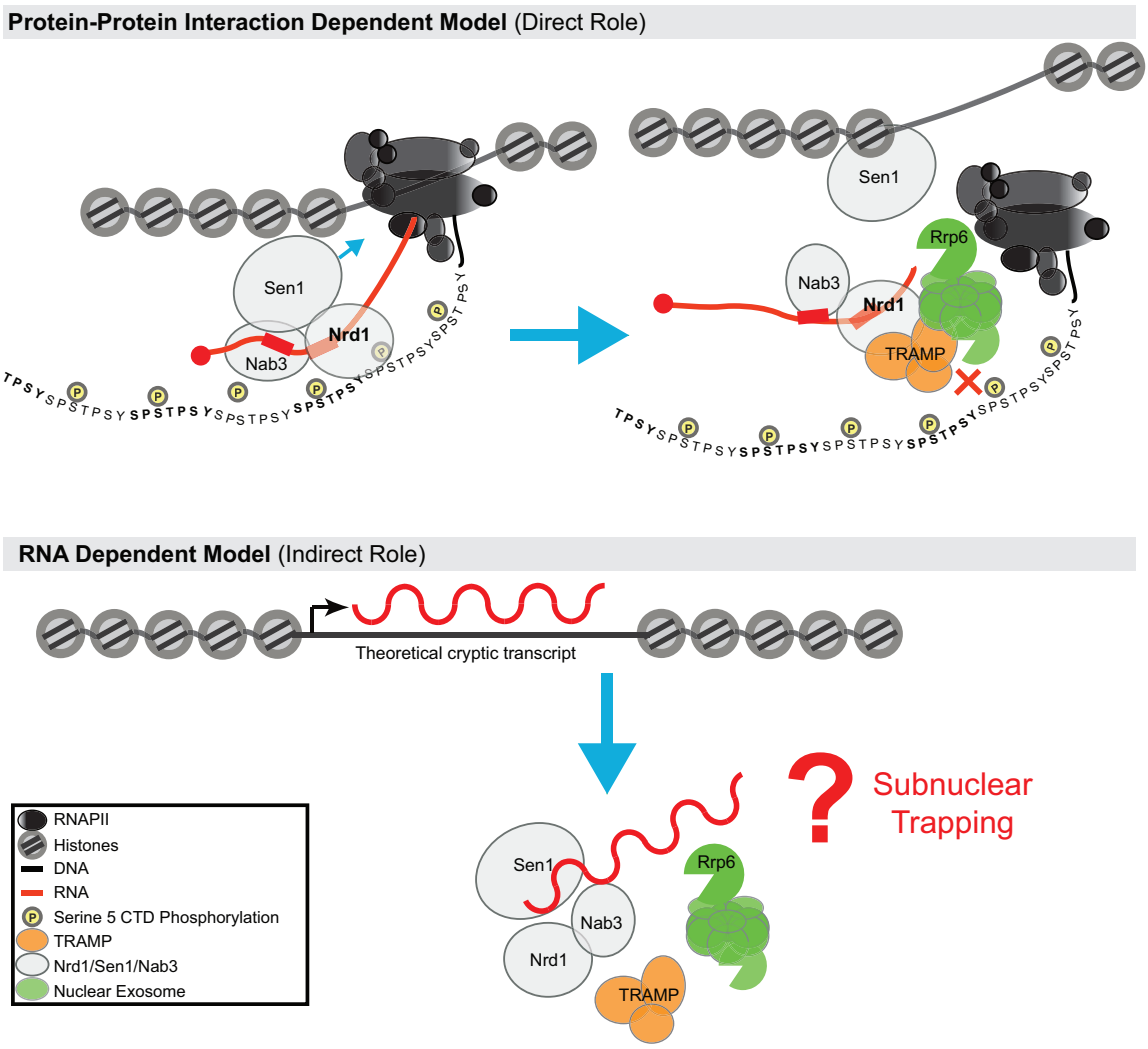

Figure 38: Two models describing how Rrp6-activity possibly modulates NNS-dependent transcription termination. Rrp6 may have a direct role in NNS-dependent termination through protein-protein interactions or an indirect role through an unidentified RNA intermediary. (Top) Nrd1 binds Ser5-P on the RNAPII CTD, and Nrd1 and Nab3 bind the nascent RNA. The mechanism by which Nrd1 causes termination is not known, but may involve an interaction with Rrp6. Sen1 unwinds RNA:DNA hybrids in a 3'-5' direction relative to the nascent RNA facilitating RNAPII termination. The TRAMP subunit Trf4 binds Nrd1, releasing it from the CTD (indicated as a red 'X'). TRAMP adds a short polyA tail to the RNA, targeting it for degradation by the exosome. Rrp6 interacts with Nrd1, likely still bound to the RNA to inhibit complete degradation of stable transcripts such as snRNA. RNAPII must then be removed from the DNA and proteins bound to both the RNA and RNA must be released through unresolved mechanisms. (Bottom) In the RNA-dependent model, a theoretical cryptic lncRNA that is usually degraded by Rrp6 may be stabilized in its absence. This trans-acting RNA may then sequester one or more proteins required for NNS-

dependent termination. This subnuclear trapping may be similar to the mechanism by which NEAT1 sequesters RNA-binding proteins in paraspeckles, effectively inhibiting their activity without decreasing their overall abundance. Nrd1 and Nab3 would be reasonable candidates for such a mechanism considering their RRM domains, or another currently unidentified protein could also be affected.

6.2. Pros and cons of a compact genome

As we have illustrated in our RNA-Seq results, it is critical that RNAPII transcription termination occurs in a limited region. We clearly show that extended transcription resulting from improper termination can effect the expression of neighboring genes (Figures 16, 17, 20). Proper regulation of termination is therefore not only required for the expression of the specific target gene but also has implications for global RNA expression levels. As discussed above, the genome of *S. cerevisiae* is extremely compact. Almost all of the genome is transcribed in some capacity, and even highly transcribed regions including snRNA genes and mRNA genes converge and overlap (examples in Figures 16-20, 22, 23, and 26). A compact genome has less room for error when regulating the location of transcription termination, but it also required less energy to replicate during cell division. It is our current understanding that genes are not as tightly packed in the more complex mammalian genomes. These genomes are much larger and require more energy to duplicate, but they also contain various *cis* and *trans* regulatory elements in non-coding genomic regions that fine tune gene expression. Altered RNAPII transcription caused by events like termination defects could potentially have a more detrimental effect on genomic expression in such a complicated system. Therefore, the additional energy required to replicate and maintain a larger, less compact genome may be a required cost of maintaining a complex, multicellular organism. However, this does not negate the requirement for proper regulation of transcription termination in mammalian genomes. Data from *S. cerevisiae* has shown that transcription

can be extended several kilobases when termination is disrupted, such as at the antisense transcript of *FMP40* when Nrd1 is depleted from the nucleus [214]. Transcription extending to such lengths would still interfere with protein coding genes in a mammalian system.

7. Role of Rtr1 in the NNS-dependent termination pathway

7.1. Rtr1 limits the co-occupancy of Nrd1/RNAPII

By comparison of RNA-Seq, ChIP-exo, and affinity purification mass spectrometry, our data clearly show that Nrd1/RNAPII co-localization is increased and NNS-dependent termination is enhanced in the absence of the atypical phosphatase Rtr1. Additionally, loss of *RTR1* is shown to have the opposite phenotype as the *RRP6* deletion at many NNS-target transcripts, and the *rtr1Δ rrp6Δ* phenocopies the *rrp6Δ* strain. These findings support the hypothesis that Rtr1 helps to fine-tune NNS-dependent termination of transcription, particularly of non-coding gene transcripts that can regulate adjacent or overlapping protein-coding genes. Furthermore, our data indicate that this role of Rtr1 is carried out directly through the NNS-dependent termination pathway and is epistatic to the effects of the *rrp6Δ*.

Our group and others have previously shown that Rtr1 is capable of removing both Tyr1-P and Ser5-P on the RNAPII CTD [8, 17], but the mechanism of Rtr1 phosphatase activity and the scope of the effects of Rtr1 on transcription are not well understood. Our findings are the first to indicate that Rtr1 plays a role in the termination of short, noncoding RNAs and to show

changes in Nrd1 occupancy that are caused by *rtr1Δ*. We note that loss of *RTR1* results in a 5'-shift of Nrd1 occupancy. Conversely, disruption of Ssu72, another Ser5 phosphatase, results in NNS-dependent terminator read-through and longer Ssu72-responsive transcripts [218, 223]. These data suggest that Rtr1 and Ssu72 play opposite roles in the regulation of NNS termination perhaps through regulation of context specific Ser5 CTD phosphorylation.

Our data further supports the hypothesis that the NNS-dependent terminator mechanism is an inefficient process requiring multiple Nrd1/Nab3-binding sites at some target genes such as *NRD1* [40, 222], and that Rtr1 plays a role in regulating its efficiency. As mentioned previously, leaky transcription attenuation at genes such as *NRD1*, *IMD2*, and *SRG1-SER3* adds a layer of regulation at these tightly controlled genes. It could be presumed that such attenuation mechanisms, which could have the ability to completely abolish gene expression by an abundance of Nrd1/Nab3, would have multiple avenues for regulation. Our findings suggest that Rrp6 and Rtr1 play opposing, yet complementary roles in fine-tuning NNS-dependent termination. These data support a model (illustrated in Figure 39) in which Rtr1 activity in WT cells limits the interaction of Nrd1 and RNAPII, likely by decreasing the number of potential Nrd1-docking sites (Ser5-P) along the CTD (Figure 39A). More Rtr1 activity therefore decreases the efficiency of NNS-dependent termination resulting in a leakier system. Without Rtr1, or in conditions in which Rtr1 protein levels are decreased, Nrd1 can bind more efficiently to RNAPII (Figure 39B). Once Nrd1 is bound to RNAPII in the ratio determined, at least in part, by Rtr1 and Kin28 (the Ser5 kinase) activity, the

interaction between Nrd1 and Rrp6 is required for proper termination (Figure 39C). Without Rrp6, the 3'-end of the transcript may be improperly processed, destabilizing the RNA, which may then be degraded by other RNases such as Dis3, resulting in decreased expression regardless of the upstream affects of Rtr1 (Figure 39D). Therefore, these two proteins, and likely others, work in concert to regulate NNS-dependent gene expression attenuation.

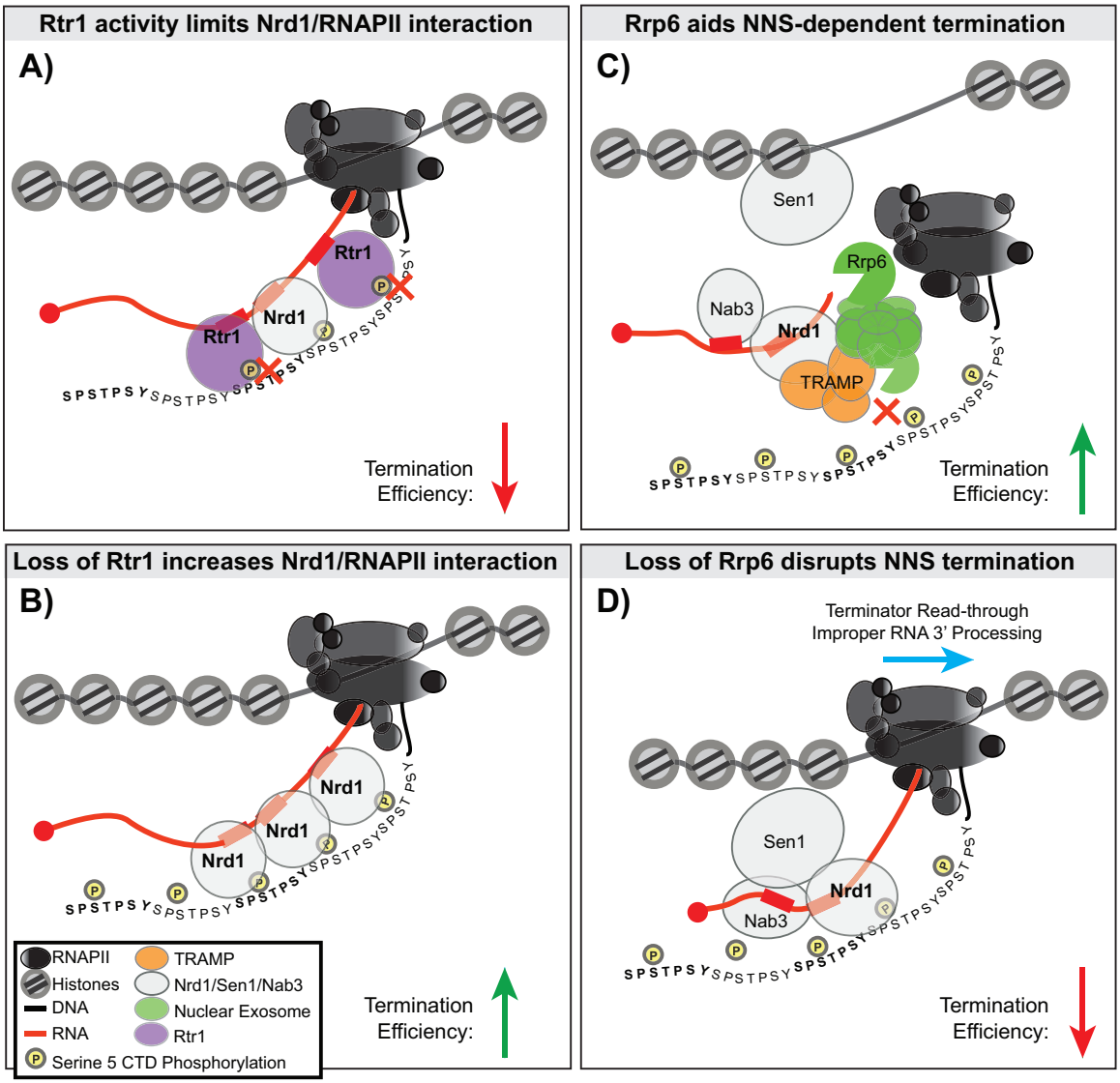

Figure 39: Representation of the effects of Rtr1 and Rrp6 on NNS-dependent transcription termination. A) Rtr1 activity in WT cells limits the interaction of Nrd1 and RNAPII, likely by decreasing the number of potential Nrd1-docking sites. B) Without Rtr1 Nrd1 can bind more efficiently to RNAPII. C) The interaction between Nrd1 and Rrp6 is required for proper termination. D) NNS-dependent termination is inefficient at certain transcripts without Rrp6, resulting in terminator read-through and improperly processed RNAs.

7.2. Comparison of Rtr1 and Ssu72 in NNS-dependent termination

These studies are the first to uncover the differences in the molecular roles of the two Ser5 phosphates Rtr1 and Ssu72. An earlier transcriptome-wide study measured the global effects on RNA of inactivation of Ssu72 using tiling microarrays [218]. The loss of Ssu72 activity in the temperature sensitive strain *SSU72-2* resulted in increased detection of noncoding RNAs transcribed from bidirectional promoters, a class of transcripts the authors termed Ssu72-restricted transcripts (SRTs). The promoters associated with these SRTs were also found to have a more relaxed chromatin structure in the absence of Ssu72 activity. Ssu72 has also been implicated in gene-looping, a model in which promoter and terminator regions of a gene are brought together through protein-protein interactions in a termination-dependent mechanism. This loop formation facilitates RNAPII transcription re-initiation and requires CTD dephosphorylation by Ssu72 and TFIIB binding of RNAPII at the terminator that then associates with the promoter [371, 372]. These findings fit well with data showing that Ssu72 occupancy at mRNA genes peaks primarily at the transcription termination site but also associates to a lesser extent at the transcription start site, independent on transcription rate and gene length [188]. Conversely, Rtr1 occupancy peaks during mid-to-late elongation during the Ser5-P to Ser2-P transition [8]. Taken together, these data support a working model in which Ssu72-mediated gene-looping promotes RNAPII transcription directionality [218].

Conversely, our data shows that loss of Rtr1 has a much larger effect on antisense noncoding transcripts that occur within the gene body of a protein-

coding gene. It is evident that noncoding transcripts originating from within a gene body or from a bidirectional promoter are regulated by different mechanisms. As mentioned previously, degradation of many noncoding transcripts from within a gene require Rrp6 activity whereas transcripts originating from a bidirectional promoter can be degraded by the core exosome and do not require Rrp6 [243]. The work described here does not address whether or not regulation of antisense transcripts is dependent on the phosphatase activity of Rtr1, but the discovery that the loss of Rtr1 enhances Nrd1 and RNAPII interaction suggests that it may be.

Furthermore, we showed that loss of *RTR1* has the opposite effect of Ssu72-inactivation at previously studied transcripts such as *IMD2* ([223] and Figure 34, 35) and *FMP40* [218]. As mentioned, *IMD2* is regulated in part by an upstream noncoding RNA that is terminated by the NNS pathway. Our data indicate that NNS-dependent termination is more efficient at this non-coding transcript in the absence of *RTR1* resulting in decreased expression of this upstream transcript and the *IMD2* gene. However, Loya *et al.* found that an *SSU72* mutant they termed Ssu72 TOV (Terminator Override) inhibited proper NNS-dependent termination at the *IMD2* upstream transcript upon loss of Ssu72 activity [223]. Likewise, our data clearly show that NNS-dependent termination of the non-coding transcript *YPL222C-A* is more efficient upon loss of *RTR1*, but tiling microarray data interrogating the effects of the Ssu72 temperature sensitive mutant *ssu72-2* shows that this transcript is extended without proper Ssu72 activity, indicating that RNAPII termination is inhibited [218]. At both of these

transcripts, NNS-dependent termination is found to require Ssu72 activity but be limited by the presence of Rtr1.

A central question remains: how do these two Ser5 phosphatases confer their distinct effects on RNAPII CTD phosphorylation and gene transcription? There are at least three possibilities. First, Rtr1- and Ssu72-specific protein-protein interactions may cause these distinct downstream effects. Second, Rtr1 and Ssu72 may not target the exact same substrate. The combinatorial nature of post-translational modifications along the CTD may result in populations of Ser5 repeats that are preferentially dephosphorylated by one or the other. For instance, it is known that Ssu72 prefers to dephosphorylate Ser5 that is in a *cis*-Ser5-P-Pro6 confirmation, isomerized by the CTD factor Ess1 [63, 373, 374]. It is still unclear what form Rtr1 prefers. The third possibility is that both Rtr1 and Ssu72 have been found to have other targets along the CTD. Rtr1 can dephosphorylate Tyr1 [17], and Ssu72 can dephosphorylate Ser7 [188, 190]. Future structural and phosphatase activity studies focused on Rtr1 may elucidate how the targeting mechanisms of these enzymes differ and what, if any, effect they may have on the regulation of RNAPII transcription.

8. Potential significance of Rbp3 ChIP-exo and Nrd1 PAR-CLIP overlap

8.1. Implications for RNA binding proteins in transcription.

This study is the first to perform high-resolution ChIP-exo analysis on RNAPII. One benefit of the ChIP-exo technique is that we can detect peaks of RNAPII occupancy adjacent to Nrd1 RNA-binding sites previously reported using PAR-

CLIP [50] (examples in Figures 16 and 21). We hypothesize that the RNAPII peaks are the result of paused polymerase in the undergoing NNS-dependent termination. The *NRD1* gene is a particularly striking example (Figure 21). At *NRD1*, there are three distinguishable Nrd1 RNA-binding sites, and each RNA-binding site is accompanied by a slightly offset peak of RNAPII. As discussed above, Nrd1 binds both the nascent RNA (via its RRM) and the RNAPII CTD (via its CID) [19, 45, 375]. Furthermore, Sen1 has been proposed to follow RNAPII along the transcribed region until Sen1 reached RNAPII triggering termination [51]. We propose that Nrd1 functions in NNS-dependent termination to anchor RNAPII to the nascent RNA near the termination site. The binding of Nrd1 to both the RNA and RNAPII may act to slow the transcribing polymerase at the required point of termination, allowing Sen1 to reach RNAPII, resulting in termination.

8.2. Future Directions

Many questions remain regarding the mechanism of NNS-dependent termination. The model proposed above in which Nrd1 slows the transcribing polymerase to allow for transcription termination would be a major step forward in our understanding of NNS-dependent termination if shown to be true. In this model, we propose that S5-P and Nrd1-binding sites within the nascent RNA are sufficient to recruit Nrd1 and slow RNAPII. This hypothesis could be tested *in vivo* by adding Nrd1-binding sites to the yeast genome in a location where transcribing RNAPII would be phosphorylated at S5. We propose that embedding Nrd1-binding sites within a 5'-UTR of a highly transcribed gene would result in

recruitment of Nrd1 and pausing of RNAPII. These outcomes could be detected using ChIP-exo to measure the localization of Nrd1-TAP and Rpb3-FLAG as described in this thesis.

8.3. Implications for strict NNS regulation by changing the expression level/nuclear localization of Rtr1 and the exosome

Many NNS-target transcripts, such as *NRD1* contain multiple Nrd1-binding sites. Our data indicate that varying the abundance of proteins such as Rtr1 and Rrp6 alters the efficiency with which NNS-dependent termination occurs. We could speculate that, by varying the rate at which Nrd1 slows RNAPII transcription at these redundant Nrd1-binding sites, the NNS-dependent termination pathway provides, not an on/off switch, but a rheostat serving to dial the efficiency of transcription termination up or down. This model may also help to describe some of the gene-specific effects we see regarding the role of Rtr1 and Rrp6 in NNS-dependent termination. Perhaps, the NNS-dependent “rheostat” is more sensitive to perturbation at some genes than others, allowing for a wider range of genome-wide effects from the regulation of NNS-dependent termination.

REFERENCES

1. Buratowski S: **Progression through the RNA polymerase II CTD cycle**. *Mol Cell* 2009, **36**(4):541-546.
2. Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F: **A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes**. *Mol Cell* 2012, **45**(2):158-170.
3. Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P: **Uniform transitions of the general RNA polymerase II transcription complex**. *Nat Struct Mol Biol* 2010, **17**(10):1272-1278.
4. Gudipati RK, Villa T, Boulay J, Libri D: **Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice**. *Nat Struct Mol Biol* 2008, **15**(8):786-794.
5. Wong KH, Jin Y, Struhl K: **TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape**. *Mol Cell* 2014, **54**(4):601-612.
6. Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ: **TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II**. *Mol Cell* 2009, **34**(3):387-393.
7. Valay JG, Simon M, Dubois MF, Bensaude O, Facca C, Faye G: **The KIN28 gene is required both for RNA polymerase II mediated**

transcription and phosphorylation of the Rpb1p CTD. *J Mol Biol* 1995, **249**(3):535-544.

8. Mosley AL, Pattenden SG, Carey M, Venkatesh S, Gilmore JM, Florens L, Workman JL, Washburn MP: **Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation**. *Mol Cell* 2009, **34**(2):168-178.
9. Smith-Kinnaman WR, Berna MJ, Hunter GO, True JD, Hsu P, Cabello GI, Fox MJ, Varani G, Mosley AL: **The interactome of the atypical phosphatase Rtr1 in Saccharomyces cerevisiae**. *Mol Biosyst* 2014, **10**(7):1730-1741.
10. Patturajan M, Conrad NK, Bregman DB, Corden JL: **Yeast carboxyl-terminal domain kinase I positively and negatively regulates RNA polymerase II carboxyl-terminal domain phosphorylation**. *J Biol Chem* 1999, **274**(39):27823-27828.
11. Tietjen JR, Zhang DW, Rodriguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X, Chapman RD, Shokat K, Keles S *et al*: **Chemical-genomic dissection of the CTD code**. *Nat Struct Mol Biol* 2010, **17**(9):1154-1161.
12. Xiang K, Nagaike T, Xiang S, Kilic T, Beh MM, Manley JL, Tong L: **Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex**. *Nature* 2010, **467**(7316):729-733.
13. Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P: **CTD tyrosine phosphorylation**

impairs termination factor recruitment to RNA polymerase II. *Science* 2012, **336**(6089):1723-1725.

14. Schreieck A, Easter AD, Etzold S, Wiederhold K, Lidschreiber M, Cramer P, Passmore LA: **RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7**. *Nat Struct Mol Biol* 2014, **21**(2):175-179.

15. Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, Hughes T, Buratowski S, Moore CL, Greenblatt J: **Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends**. *J Biol Chem* 2003, **278**(35):33000-33010.

16. Nedea E, Nalbant D, Xia D, Theoharis NT, Suter B, Richardson CJ, Tatchell K, Kislinger T, Greenblatt JF, Nagy PL: **The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes**. *Mol Cell* 2008, **29**(5):577-587.

17. Hsu PL, Yang F, Smith-Kinnaman W, Yang W, Song JE, Mosley AL, Varani G: **Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD**. *J Mol Biol* 2014, **426**(16):2970-2981.

18. Rondon AGM, Hannah E.; Proudfoot, Nick J.: **Terminating Transcription in Yeast: whether to be a 'nerd' or a 'rat'**. *Nat Struct Mol Biol* 2008, **15**(8):775-776.

19. Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A: **The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain**. *Nat Struct Mol Biol* 2008, **15**(8):795-804.

20. Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S: **Distinct pathways for snoRNA and mRNA termination**. *Mol Cell* 2006, **24**(5):723-734.

21. Kuehner JN, Pearson EL, Moore C: **Unravelling the means to an end: RNA polymerase II transcription termination**. *Nat Rev Mol Cell Biol* 2011, **12**(5):283-294.

22. Kyburz A, Sadowski M, Dichtl B, Keller W: **The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation**. *Nucleic Acids Res* 2003, **31**(14):3936-3945.

23. Ezeokonkwo C, Zhelkovsky A, Lee R, Bohm A, Moore CL: **A flexible linker region in Fip1 is needed for efficient mRNA polyadenylation**. *Rna* 2011, **17**(4):652-664.

24. Vo LT, Minet M, Schmitter JM, Lacroute F, Wyers F: **Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA**. *Mol Cell Biol* 2001, **21**(24):8346-8356.

25. Danckwardt SH, Matthias W.; Kulozik, Andreas: **3' end mRNA processing; molecular mechanisms and implications for health and disease**. *Embo J* 2008, **27**:482-498.

26. Ohnacker M BS, Preker PJ, Keller W.: **The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex**. *Embo J* 2000, **19**(1):37-47.

27. Buchert M, Papin M, Bonnans C, Darido C, Raye WS, Garambois V, Pelegrin A, Bourgaux JF, Pannequin J, Joubert D *et al*: **Symplekin promotes tumorigenicity by up-regulating claudin-2 expression**. *Proc Natl Acad Sci U S A* 2010, **107**(6):2628-2633.

28. Das BG, Zijian; Russo, Patrick; Chartrand, Pascal; Sherman, Fred: **The role of nuclear cap binding protein Cdc1p of yeast in mRNA termination and degradation**. *Mol Cell Biol* 2000, **20**(8):2827-2838.

29. Holbein S, Scola S, Loll B, Dichtl BS, Hubner W, Meinhart A, Dichtl B: **The P-loop domain of yeast Clp1 mediates interactions between CF IA and CPF factors in pre-mRNA 3' end formation**. *PLoS One* 2011, **6**(12):e29139.

30. Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S: **The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II**. *Nature* 2004, **432**(7016):517-522.

31. Zhao J, Kessler M, Helmling S, O'Connor JP, Moore C: **Pta1, a component of yeast CF II, is required for both cleavage and poly(A) addition of mRNA precursor**. *Mol Cell Biol* 1999, **19**(11):7733-7740.

32. Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G: **Cooperative interaction of**

transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 2010, **17**(10):1195-1201.

33. Xiang K, Tong L, Manley JL: **Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery**. *Mol Cell Biol* 2014, **34**(11):1894-1910.

34. Gromak N, West S, Proudfoot NJ: **Pause sites promote transcriptional termination of mammalian RNA polymerase II**. *Mol Cell Biol* 2006, **26**(10):3986-3996.

35. Glover-Cutter K, Kim S, Espinosa J, Bentley DL: **RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes**. *Nat Struct Mol Biol* 2008, **15**(1):71-78.

36. Anderson JT: **RNA turnover: unexpected consequences of being tailed**. *Curr Biol* 2005, **15**(16):R635-638.

37. Nag A, Narsinh K, Martinson HG: **The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase**. *Nat Struct Mol Biol* 2007, **14**(7):662-669.

38. Proudfoot N: **New perspectives on connecting messenger RNA 3' end formation to transcription**. *Current opinion in cell biology* 2004, **16**(3):272-278.

39. Buratowski S: **Connections between mRNA 3' end processing and transcription termination**. *Current opinion in cell biology* 2005, **17**(3):257-261.

40. Carroll KL, Ghirlando R, Ames JM, Corden JL: **Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements**. *Rna* 2007, **13**(3):361-373.

41. Arigo JT, Eyler DE, Carroll KL, Corden JL: **Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3**. *Mol Cell* 2006, **23**(6):841-851.

42. Carroll KL, Pradhan DA, Granek JA, Clarke ND, Corden JL: **Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts**. *Mol Cell Biol* 2004, **24**(14):6241-6252.

43. Steinmetz EJ, Conrad NK, Brow DA, Corden JL: **RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts**. *Nature* 2001, **413**(6853):327-331.

44. Webb S, Hector, R.D., Kudla, G., Granneman, S.: **PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast**. *Genome Biol* 2014, **15**(1):R8.

45. Heo DH, Yoo I, Kong J, Lidschreiber M, Mayer A, Choi BY, Hahn Y, Cramer P, Buratowski S, Kim M: **The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome**. *J Biol Chem* 2013, **288**(51):36676-36690.

46. Meinhart A, Cramer P: **Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors**. *Nature* 2004, **430**(6996):223-226.

47. Vasiljeva L, Kim M, Terzi N, Soares LM, Buratowski S: **Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin**. *Mol Cell* 2008, **29**(3):313-323.

48. Conrad NK, Wilson SM, Steinmetz EJ, Patturajan M, Brow DA, Swanson MS, Corden JL: **A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II**. *Genetics* 2000, **154**(2):557-571.

49. Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ, Corden JL: **Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing**. *Rna* 2011, **17**(11):2011-2025.

50. Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ, Corden JL: **Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1**. *PLoS Genet* 2011, **7**(10):e1002329.

51. Hazelbaker DZ, Marquardt S, Wlotzka W, Buratowski S: **Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination**. *Mol Cell* 2013, **49**(1):55-66.

52. Brow DA: **Sen-sing RNA terminators**. *Mol Cell* 2011, **42**(6):717-718.

53. Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ, Brow DA: **Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase**. *Mol Cell* 2006, **24**(5):735-746.

54. Skourti-Stathaki K, Proudfoot NJ, Gromak N: **Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination**. *Mol Cell* 2011, **42**(6):794-805.

55. Porrua O, Libri D: **A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast**. *Nat Struct Mol Biol* 2013, **20**(7):884-891.

56. Honorine R, Mosrin-Huaman C, Hervouet-Coste N, Libri D, Rahmouni AR: **Nuclear mRNA quality control in yeast is mediated by Nrd1 co-transcriptional recruitment, as revealed by the targeting of Rho-induced aberrant transcripts**. *Nucleic Acids Res* 2011, **39**(7):2809-2820.

57. Lykke-Andersen S, Jensen TH: **Overlapping pathways dictate termination of RNA polymerase II transcription**. *Biochimie* 2007, **89**(10):1177-1182.

58. Grzechnik P, Kufel J: **Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast**. *Mol Cell* 2008, **32**(2):247-258.

59. Stuparevic I, Mosrin-Huaman C, Hervouet-Coste N, Remenaric M, Rahmouni AR: **Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p**

in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP) in yeast. *J Biol Chem* 2013, **288**(44):31816-31829.

60. Wlotzka W, Kudla G, Granneman S, Tollervey D: **The nuclear RNA polymerase II surveillance system targets polymerase III transcripts**. *Embo J* 2011, **30**(9):1790-1803.

61. Tudek A, Porrua O, Kabzinski T, Lidschreiber M, Kubicek K, Fortova A, Lacroute F, Vanacova S, Cramer P, Stefl R *et al*: **Molecular basis for coordinating transcription termination with noncoding RNA degradation**. *Mol Cell* 2014, **55**(3):467-481.

62. Grzechnik P, Gdula MR, Proudfoot NJ: **Pcf11 orchestrates transcription termination pathways in yeast**. *Genes Dev* 2015, **29**(8):849-861.

63. Singh N, Ma Z, Gemmill T, Wu X, Defiglio H, Rossettini A, Rabeler C, Beane O, Morse RH, Palumbo MJ *et al*: **The Ess1 prolyl isomerase is required for transcription termination of small noncoding RNAs via the Nrd1 pathway**. *Mol Cell* 2009, **36**(2):255-266.

64. Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D: **The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases**. *Cell* 1997, **91**(4):457-466.

65. Bonneau F, Basquin J, Ebert J, Lorentzen E, Conti E: **The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation**. *Cell* 2009, **139**(3):547-559.

66. Schneider C, Tollervey D: **Threading the barrel of the RNA exosome**. *Trends Biochem Sci* 2013, **38**(10):485-493.

67. Makino DL, Baumgartner M, Conti E: **Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex**. *Nature* 2013, **495**(7439):70-75.

68. Makino DL, Halbach F, Conti E: **The RNA exosome and proteasome: common principles of degradation control**. *Nat Rev Mol Cell Biol* 2013, **14**(10):654-660.

69. van Dijk EL, Schilders G, Pruijn GJ: **Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways**. *Rna* 2007, **13**(7):1027-1035.

70. Vasiljeva L, Buratowski S: **Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts**. *Mol Cell* 2006, **21**(2):239-248.

71. Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Regnault B, Devaux F, Namane A, Seraphin B *et al*: **Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase**. *Cell* 2005, **121**(5):725-737.

72. Neil H, Malabat C, d'Aubenton-Carafa Y, Xu Z, Steinmetz LM, Jacquier A: **Widespread bidirectional promoters are the major source of cryptic transcripts in yeast**. *Nature* 2009, **457**(7232):1038-1042.

73. Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Munster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM: **Bidirectional promoters**

generate pervasive transcription in yeast. *Nature* 2009, **457**(7232):1033-1037.

74. Porrua O, Hobor F, Boulay J, Kubicek K, D'Aubenton-Carafa Y, Gudipati RK, Stefl R, Libri D: **In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination**. *Embo J* 2012, **31**(19):3935-3948.

75. Davis CA, Ares M, Jr.: **Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae**. *Proc Natl Acad Sci U S A* 2006, **103**(9):3262-3267.

76. Dichtl B, Blank D, Ohnacker M, Friedlein A, Roeder D, Langen H, Keller W: **A role for SSU72 in balancing RNA polymerase II transcription elongation and termination**. *Mol Cell* 2002, **10**(5):1139-1150.

77. Schmid M, Poulsen MB, Olszewski P, Pelechano V, Saguez C, Gupta I, Steinmetz LM, Moore C, Jensen TH: **Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins**. *Mol Cell* 2012, **47**(2):267-280.

78. Grenier St-Sauveur V, Soucek S, Corbett AH, Bachand F: **Poly(A) tail-mediated gene regulation by opposing roles of Nab2 and Pab2 nuclear poly(A)-binding proteins in pre-mRNA decay**. *Mol Cell Biol* 2013, **33**(23):4718-4731.

79. Gudipati RK, Xu Z, Lebreton A, Seraphin B, Steinmetz LM, Jacquier A, Libri D: **Extensive degradation of RNA precursors by the exosome in wild-type cells**. *Mol Cell* 2012, **48**(3):409-421.

80. Schneider C, Kudla G, Wlotzka W, Tuck A, Tollervey D: **Transcriptome-wide analysis of exosome targets**. *Mol Cell* 2012, **48**(3):422-433.

81. Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, Chu A, Davis RW, Steinmetz LM, Primig M: **Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6**. *Proc Natl Acad Sci U S A* 2011, **108**(3):1058-1063.

82. Frenk S, Oxley D, Houseley J: **The nuclear exosome is active and important during budding yeast meiosis**. *PLoS One* 2014, **9**(9):e107648.

83. Brouwer R, Pruijn GJ, van Venrooij WJ: **The human exosome: an autoantigenic complex of exoribonucleases in myositis and scleroderma**. *Arthritis Res* 2001, **3**(2):102-106.

84. Wan J, Yourshaw M, Mamsa H, Rudnik-Schoneborn S, Menezes MP, Hong JE, Leong DW, Senderek J, Salman MS, Chitayat D *et al*: **Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration**. *Nat Genet* 2012, **44**(6):704-708.

85. Boczonadi V, Muller JS, Pyle A, Munkley J, Dor T, Quartararo J, Ferrero I, Karcagi V, Giunta M, Polvikoski T *et al*: **EXOSC8 mutations alter mRNA**

metabolism and cause hypomyelination with spinal muscular atrophy and cerebellar hypoplasia. *Nat Commun* 2014, **5**:4287.

86. Richard P, Feng S, Manley JL: **A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage**. *Genes Dev* 2013, **27**(20):2227-2232.

87. Dziembowski A, Lorentzen E, Conti E, Seraphin B: **A single subunit, Dis3, is essentially responsible for yeast exosome core activity**. *Nat Struct Mol Biol* 2007, **14**(1):15-22.

88. Lorentzen E, Walter P, Fribourg S, Evguenieva-Hackenberg E, Klug G, Conti E: **The archaeal exosome core is a hexameric ring structure with three catalytic subunits**. *Nat Struct Mol Biol* 2005, **12**(7):575-581.

89. Buttner K, Wenig K, Hopfner KP: **Structural framework for the mechanism of archaeal exosomes in RNA processing**. *Mol Cell* 2005, **20**(3):461-471.

90. Liu Q, Greimann JC, Lima CD: **Reconstitution, activities, and structure of the eukaryotic RNA exosome**. *Cell* 2006, **127**(6):1223-1237.

91. Schaeffer D, Tsanova B, Barbas A, Reis FP, Dastidar EG, Sanchez-Rotunno M, Arraiano CM, van Hoof A: **The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities**. *Nat Struct Mol Biol* 2009, **16**(1):56-62.

92. Wang HW, Wang J, Ding F, Callahan K, Bratkowski MA, Butler JS, Nogales E, Ke A: **Architecture of the yeast Rrp44 exosome complex**

suggests routes of RNA recruitment for 3' end processing. *Proc Natl Acad Sci U S A* 2007, **104**(43):16844-16849.

93. Malet H, Topf M, Clare DK, Ebert J, Bonneau F, Basquin J, Drazkowska K, Tomecki R, Dziembowski A, Conti E *et al*: **RNA channelling by the eukaryotic exosome**. *EMBO Rep* 2010, **11**(12):936-942.

94. Staals RH, Bronkhorst AW, Schilders G, Slomovic S, Schuster G, Heck AJ, Raijmakers R, Pruijn GJ: **Dis3-like 1: a novel exoribonuclease associated with the human exosome**. *Embo J* 2010, **29**(14):2358-2367.

95. Tomecki R, Kristiansen MS, Lykke-Andersen S, Chlebowski A, Larsen KM, Szczesny RJ, Drazkowska K, Pastula A, Andersen JS, Stepien PP *et al*: **The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L**. *Embo J* 2010, **29**(14):2342-2357.

96. Briggs MW, Burkard KT, Butler JS: **Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation**. *J Biol Chem* 1998, **273**(21):13255-13263.

97. Mitchell P, Petfalski E, Houalla R, Podtelejnikov A, Mann M, Tollervey D: **Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs**. *Mol Cell Biol* 2003, **23**(19):6982-6992.

98. Alderuccio F, Chan EK, Tan EM: **Molecular characterization of an autoantigen of PM-Scl in the polymyositis/scleroderma overlap syndrome: a unique and complete human cDNA encoding an**

apparent 75-kD acidic protein of the nucleolar complex. *J Exp Med* 1991, **173**(4):941-952.

99. Bluthner M, Bautz FA: **Cloning and characterization of the cDNA coding for a polymyositis-scleroderma overlap syndrome-related nucleolar 100-kD protein**. *J Exp Med* 1992, **176**(4):973-980.

100. Ge Q, Frank MB, O'Brien C, Targoff IN: **Cloning of a complementary DNA coding for the 100-kD antigenic protein of the PM-Scl autoantigen**. *J Clin Invest* 1992, **90**(2):559-570.

101. Yang XF, Wu CJ, Chen L, Alyea EP, Canning C, Kantoff P, Soiffer RJ, Dranoff G, Ritz J: **CML28 is a broadly immunogenic antigen, which is overexpressed in tumor cells**. *Cancer Res* 2002, **62**(19):5517-5522.

102. Eggens VR, Barth PG, Niermeijer JM, Berg JN, Darin N, Dixit A, Fluss J, Foulds N, Fowler D, Hortobagyi T *et al*: **EXOSC3 mutations in pontocerebellar hypoplasia type 1: novel mutations and genotype-phenotype correlations**. *Orphanet J Rare Dis* 2014, **9**:23.

103. Weissbach S, Langer C, Puppe B, Nedeva T, Bach E, Kull M, Bargou R, Einsele H, Rosenwald A, Knop S *et al*: **The molecular spectrum and clinical impact of DIS3 mutations in multiple myeloma**. *Br J Haematol* 2015, **169**(1):57-70.

104. Tomecki R, Drazkowska K, Kucinski I, Stodus K, Szczesny RJ, Gruchota J, Owczarek EP, Kalisiak K, Dziembowski A: **Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA**

metabolism and suggest hDIS3 PIN domain as a potential drug target. *Nucleic Acids Res* 2014, **42**(2):1270-1290.

105. Morris MR, Astuti D, Maher ER: **Perlman syndrome: overgrowth, Wilms tumor predisposition and DIS3L2**. *Am J Med Genet C Semin Med Genet* 2013, **163C**(2):106-113.

106. Schilders G, Egberts WV, Raijmakers R, Pruijn GJ: **C1D is a major autoantibody target in patients with the polymyositis-scleroderma overlap syndrome**. *Arthritis Rheum* 2007, **56**(7):2449-2454.

107. Noguchi E, Hayashi N, Azuma Y, Seki T, Nakamura M, Nakashima N, Yanagida M, He X, Mueller U, Sazer S *et al*: **Dis3, implicated in mitotic control, binds directly to Ran and enhances the GEF activity of RCC1**. *Embo J* 1996, **15**(20):5595-5605.

108. Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D, Mitchell P: **The yeast exosome and human PM-Scl are related complexes of 3' --> 5' exonucleases**. *Genes Dev* 1999, **13**(16):2148-2158.

109. Wasmuth EV, Lima CD: **Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel**. *Mol Cell* 2012, **48**(1):133-144.

110. Schneider C, Leung E, Brown J, Tollervey D: **The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome**. *Nucleic Acids Res* 2009, **37**(4):1127-1140.

111. Drazkowska K, Tomecki R, Stodus K, Kowalska K, Czarnocki-Cieciura M, Dziembowski A: **The RNA exosome complex central channel controls both exonuclease and endonuclease Dis3 activities in vivo and in vitro**. *Nucleic Acids Res* 2013, **41**(6):3845-3858.

112. Lebreton A, Tomecki R, Dziembowski A, Seraphin B: **Endonucleolytic RNA cleavage by a eukaryotic exosome**. *Nature* 2008, **456**(7224):993-996.

113. Lorentzen E, Basquin J, Tomecki R, Dziembowski A, Conti E: **Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family**. *Mol Cell* 2008, **29**(6):717-728.

114. Liu JJ, Bratkowski MA, Liu X, Niu CY, Ke A, Wang HW: **Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM**. *Nat Struct Mol Biol* 2014, **21**(1):95-102.

115. Yang W, Lee JY, Nowotny M: **Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity**. *Mol Cell* 2006, **22**(1):5-13.

116. Callahan KP, Butler JS: **Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p**. *Nucleic Acids Res* 2008, **36**(21):6645-6655.

117. Midtgaard SF, Assenholt J, Jonstrup AT, Van LB, Jensen TH, Brodersen DE: **Structure of the nuclear exosome component Rrp6p reveals an**

interplay between the active site and the HRDC domain. *Proc Natl Acad Sci U S A* 2006, **103**(32):11898-11903.

118. Januszyk K, Liu Q, Lima CD: **Activities of human RRP6 and structure of the human RRP6 catalytic domain**. *Rna* 2011, **17**(8):1566-1577.

119. Wasmuth EV, Januszyk K, Lima CD: **Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA**. *Nature* 2014, **511**(7510):435-439.

120. Halbach F, Reichelt P, Rode M, Conti E: **The yeast ski complex: crystal structure and RNA channeling to the exosome complex**. *Cell* 2013, **154**(4):814-826.

121. Schmidt K, Butler JS: **Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity**. *Wiley Interdiscip Rev RNA* 2013, **4**(2):217-231.

122. LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D: **RNA degradation by the exosome is promoted by a nuclear polyadenylation complex**. *Cell* 2005, **121**(5):713-724.

123. Callahan KP, Butler JS: **TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6**. *J Biol Chem* 2010, **285**(6):3540-3547.

124. Kish-Trier E, Hill CP: **Structural biology of the proteasome**. *Annu Rev Biophys* 2013, **42**:29-49.

125. Burkard KT, Butler JS: **A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p**. *Mol Cell Biol* 2000, **20**(2):604-616.

126. Stead JA, Costello JL, Livingstone MJ, Mitchell P: **The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein**. *Nucleic Acids Res* 2007, **35**(16):5556-5567.

127. Feigenbutz M, Garland W, Turner M, Mitchell P: **The exosome cofactor Rrp47 is critical for the stability and normal expression of its associated exoribonuclease Rrp6 in Saccharomyces cerevisiae**. *PLoS One* 2013, **8**(11):e80752.

128. Garland W, Feigenbutz M, Turner M, Mitchell P: **Rrp47 functions in RNA surveillance and stable RNA processing when divorced from the exoribonuclease and exosome-binding domains of Rrp6**. *Rna* 2013, **19**(12):1659-1668.

129. Feigenbutz M, Jones R, Besong TM, Harding SE, Mitchell P: **Assembly of the yeast exoribonuclease Rrp6 with its associated cofactor Rrp47 occurs in the nucleus and is critical for the controlled expression of Rrp47**. *J Biol Chem* 2013, **288**(22):15959-15970.

130. Schuch B, Feigenbutz M, Makino DL, Falk S, Basquin C, Mitchell P, Conti E: **The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase**. *Embo J* 2014, **33**(23):2829-2846.

131. Chen CY, Gherzi R, Ong SE, Chan EL, Raijmakers R, Pruijn GJ, Stoecklin G, Moroni C, Mann M, Karin M: **AU binding proteins recruit the exosome to degrade ARE-containing mRNAs**. *Cell* 2001, **107**(4):451-464.

132. Matsumoto-Taniura N, Pirollet F, Monroe R, Gerace L, Westendorf JM: **Identification of novel M phase phosphoproteins by expression cloning**. *Mol Biol Cell* 1996, **7**(9):1455-1469.

133. Schilders G, Raijmakers R, Raats JM, Pruijn GJ: **MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation**. *Nucleic Acids Res* 2005, **33**(21):6795-6804.

134. Milligan L, Decourty L, Saveanu C, Rappsilber J, Ceulemans H, Jacquier A, Tollervey D: **A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts**. *Mol Cell Biol* 2008, **28**(17):5446-5457.

135. de la Cruz J, Kressler D, Tollervey D, Linder P: **Dob1p (Mtr4p) is a putative ATP-dependent RNA helicase required for the 3' end formation of 5.8S rRNA in Saccharomyces cerevisiae**. *Embo J* 1998, **17**(4):1128-1140.

136. Bernstein J, Patterson DN, Wilson GM, Toth EA: **Characterization of the essential activities of Saccharomyces cerevisiae Mtr4p, a 3'->5' helicase partner of the nuclear exosome**. *J Biol Chem* 2008, **283**(8):4930-4942.

137. Jackson RN, Klauer AA, Hintze BJ, Robinson H, van Hoof A, Johnson SJ: **The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing**. *Embo J* 2010, **29**(13):2205-2216.

138. Weir JR, Bonneau F, Hentschel J, Conti E: **Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance**. *Proc Natl Acad Sci U S A* 2010, **107**(27):12139-12144.

139. Taylor LL, Jackson RN, Rexhepaj M, King AK, Lott LK, van Hoof A, Johnson SJ: **The Mtr4 ratchet helix and arch domain both function to promote RNA unwinding**. *Nucleic Acids Res* 2014, **42**(22):13861-13872.

140. Bernstein J, Ballin JD, Patterson DN, Wilson GM, Toth EA: **Unique properties of the Mtr4p-poly(A) complex suggest a role in substrate targeting**. *Biochemistry* 2010, **49**(49):10357-10370.

141. Vanacova S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W: **A new yeast poly(A) polymerase complex involved in RNA quality control**. *PLoS Biol* 2005, **3**(6):e189.

142. Kadaba S, Krueger A, Trice T, Krecic AM, Hinnebusch AG, Anderson J: **Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae**. *Genes Dev* 2004, **18**(11):1227-1240.

143. Wang X, Jia H, Jankowsky E, Anderson JT: **Degradation of hypomodified tRNA(iMet) in vivo involves RNA-dependent ATPase activity of the DExH helicase Mtr4p**. *Rna* 2008, **14**(1):107-116.

144. Jenks MH, Reines D: **Dissection of the molecular basis of mycophenolate resistance in Saccharomyces cerevisiae**. *Yeast* 2005, **22**(15):1181-1190.

145. Steinmetz EJ, Ng SB, Cloute JP, Brow DA: **cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II**. *Mol Cell Biol* 2006, **26**(7):2688-2696.

146. Kopcewicz KA, O'Rourke TW, Reines D: **Metabolic regulation of IMD2 transcription and an unusual DNA element that generates short transcripts**. *Mol Cell Biol* 2007, **27**(8):2821-2829.

147. Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D: **Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance**. *Mol Cell* 2006, **23**(6):853-864.

148. Mosrin-Huaman C, Honorine R, Rahmouni AR: **Expression of bacterial Rho factor in yeast identifies new factors involved in the functional interplay between transcription and mRNP biogenesis**. *Mol Cell Biol* 2009, **29**(15):4033-4044.

149. Fasken MB, Laribee RN, Corbett AH: **Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rrp6 independent of Nrd1**. *PLoS Genet* 2015, **11**(3):e1005044.

150. Wagschal A, Rousset E, Basavarajaiah P, Contreras X, Harwig A, Laurent-Chabalier S, Nakamura M, Chen X, Zhang K, Meziane O *et al*: **Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce**

premature termination of transcription by RNAPII. *Cell* 2012, **150**(6):1147-1157.

151. Macias S, Cordiner RA, Caceres JF: **Cellular functions of the microprocessor**. *Biochem Soc Trans* 2013, **41**(4):838-843.

152. Thiebaut M, Colin J, Neil H, Jacquier A, Seraphin B, Lacroute F, Libri D: **Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in S. cerevisiae**. *Mol Cell* 2008, **31**(5):671-682.

153. Colin J, Libri D, Porrua O: **Cryptic transcription and early termination in the control of gene expression**. *Genet Res Int* 2011, **2011**:653494.

154. Mitchell P, Petfalski E, Tollervey D: **The 3' end of yeast 5.8S rRNA is generated by an exonuclease processing mechanism**. *Genes Dev* 1996, **10**(4):502-513.

155. Schilders G, van Dijk E, Pruijn GJ: **C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing**. *Nucleic Acids Res* 2007, **35**(8):2564-2572.

156. Allmang C, Mitchell P, Petfalski E, Tollervey D: **Degradation of ribosomal RNA precursors by the exosome**. *Nucleic Acids Res* 2000, **28**(8):1684-1691.

157. Kadaba S, Wang X, Anderson JT: **Nuclear RNA surveillance in Saccharomyces cerevisiae: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA**. *Rna* 2006, **12**(3):508-521.

158. Huang Y, Bayfield MA, Intine RV, Maraia RJ: **Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation**. *Nat Struct Mol Biol* 2006, **13**(7):611-618.

159. Schneider C, Anderson JT, Tollervey D: **The exosome subunit Rrp44 plays a direct role in RNA substrate recognition**. *Mol Cell* 2007, **27**(2):324-331.

160. Caffarelli E, Arese M, Santoro B, Fragapane P, Bozzoni I: **In vitro study of processing of the intron-encoded U16 small nucleolar RNA in Xenopus laevis**. *Mol Cell Biol* 1994, **14**(5):2966-2974.

161. Cecconi F, Mariottini P, Amaldi F: **The Xenopus intron-encoded U17 snoRNA is produced by exonucleolytic processing of its precursor in oocytes**. *Nucleic Acids Res* 1995, **23**(22):4670-4676.

162. Cavaille J, Bachellerie JP: **Processing of fibrillarin-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by the common stem-box terminal structure**. *Biochimie* 1996, **78**(6):443-456.

163. Larimer FW, Hsu CL, Maupin MK, Stevens A: **Characterization of the XRN1 gene encoding a 5'-->3' exoribonuclease: sequence data and analysis of disparate protein and mRNA levels of gene-disrupted yeast cells**. *Gene* 1992, **120**(1):51-57.

164. Petfalski E, Dandekar T, Henry Y, Tollervey D: **Processing of the precursors to small nucleolar RNAs and rRNAs requires common components**. *Mol Cell Biol* 1998, **18**(3):1181-1189.

165. Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D: **Functions of the exosome in rRNA, snoRNA and snRNA synthesis**. *Embo J* 1999, **18**(19):5399-5410.

166. van Hoof A, Lennertz P, Parker R: **Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs**. *Mol Cell Biol* 2000, **20**(2):441-452.

167. Lemay JF, D'Amours A, Lemieux C, Lackner DH, St-Sauveur VG, Bahler J, Bachand F: **The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs**. *Mol Cell* 2010, **37**(1):34-45.

168. Martens JA, Laprade L, Winston F: **Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene**. *Nature* 2004, **429**(6991):571-574.

169. Martens JA, Wu PY, Winston F: **Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae**. *Genes Dev* 2005, **19**(22):2695-2704.

170. Kuehner JN, Brow DA: **Regulation of a eukaryotic gene by GTP-dependent start site selection and transcription attenuation**. *Mol Cell* 2008, **31**(2):201-211.

171. Arigo JT, Carroll KL, Ames JM, Corden JL: **Regulation of yeast NRD1 expression by premature transcription termination**. *Mol Cell* 2006, **21**(5):641-651.

172. Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu ZP *et al*: **RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity**. *Cell* 2015, **161**(4):774-789.

173. Wang KC, Chang HY: **Molecular mechanisms of long noncoding RNAs**. *Mol Cell* 2011, **43**(6):904-914.

174. Wilusz JE, Sunwoo H, Spector DL: **Long noncoding RNAs: functional surprises from the RNA world**. *Genes Dev* 2009, **23**(13):1494-1504.

175. Nagano T, Fraser P: **No-nonsense functions for long noncoding RNAs**. *Cell* 2011, **145**(2):178-181.

176. Berretta J, Morillon A: **Pervasive transcription constitutes a new level of eukaryotic genome regulation**. *EMBO Rep* 2009, **10**(9):973-982.

177. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP *et al*: **Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals**. *Nature* 2009, **458**(7235):223-227.

178. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A *et al*: **Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression**. *Proc Natl Acad Sci U S A* 2009, **106**(28):11667-11672.

179. Galipon J, Miki A, Oda A, Inada T, Ohta K: **Stress-induced lncRNAs evade nuclear degradation and enter the translational machinery**. *Genes Cells* 2013, **18**(5):353-368.

180. Geisler S, Lojek L, Khalil AM, Baker KE, Coller J: **Decapping of long noncoding RNAs regulates inducible genes**. *Mol Cell* 2012, **45**(3):279-291.

181. Kim M, Suh H, Cho EJ, Buratowski S: **Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7**. *J Biol Chem* 2009, **284**(39):26421-26426.

182. Rosado-Lugo JD, Hampsey M: **The Ssu72 phosphatase mediates the RNA polymerase II initiation-elongation transition**. *J Biol Chem* 2014, **289**(49):33916-33926.

183. Egloff S, Zaborowska J, Laitem C, Kiss T, Murphy S: **Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes**. *Mol Cell* 2012, **45**(1):111-122.

184. Ni Z, Xu C, Guo X, Hunter GO, Kuznetsova OV, Tempel W, Marcon E, Zhong G, Guo H, Kuo WH *et al*: **RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation**. *Nat Struct Mol Biol* 2014, **21**(8):686-695.

185. Gibney PA, Fries T, Bailer SM, Morano KA: **Rtr1 is the Saccharomyces cerevisiae homolog of a novel family of RNA polymerase II-binding proteins**. *Eukaryot Cell* 2008, **7**(6):938-948.

186. Ni Z, Olsen JB, Guo X, Zhong G, Ruan ED, Marcon E, Young P, Guo H, Li J, Moffat J *et al*: **Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B**. *Transcription* 2011, **2**(5):237-242.

187. Xiang K, Manley JL, Tong L: **The yeast regulator of transcription protein Rtr1 lacks an active site and phosphatase activity**. *Nat Commun* 2012, **3**:946.

188. Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ: **Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination**. *J Biol Chem* 2012, **287**(11):8541-8551.

189. Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M: **Ssu72 Is an RNA polymerase II CTD phosphatase**. *Mol Cell* 2004, **14**(3):387-394.

190. Xiang K, Manley JL, Tong L: **An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72**. *Genes Dev* 2012, **26**(20):2265-2270.

191. Wani S, Yuda M, Fujiwara Y, Yamamoto M, Harada F, Ohkuma Y, Hirose Y: **Vertebrate Ssu72 regulates and coordinates 3'-end formation of RNAs transcribed by RNA polymerase II**. *PLoS One* 2014, **9**(8):e106040.

192. Schwer B, Sanchez AM, Shuman S: **Punctuation and syntax of the RNA polymerase II CTD code in fission yeast**. *Proc Natl Acad Sci U S A* 2012, **109**(44):18024-18029.

193. Steinmetz EJ, Brow DA: **Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription**. *Mol Cell Biol* 2003, **23**(18):6339-6349.

194. Ghazy MA, He X, Singh BN, Hampsey M, Moore C: **The essential N terminus of the Pta1 scaffold protein is required for snoRNA transcription termination and Ssu72 function but is dispensable for pre-mRNA 3'-end processing**. *Mol Cell Biol* 2009, **29**(8):2296-2307.

195. Luo Y, Yogesha SD, Cannon JR, Yan W, Ellington AD, Brodbelt JS, Zhang Y: **Novel Modifications on C-terminal Domain of RNA Polymerase II Can Fine-tune the Phosphatase Activity of Ssu72**. *ACS Chem Biol* 2013, **8**(9):2042-2052.

196. Allepuz-Fuster P, Martinez-Fernandez V, Garrido-Godino AI, Alonso-Aguado S, Hanes SD, Navarro F, Calvo O: **Rpb4/7 facilitates RNA polymerase II CTD dephosphorylation**. *Nucleic Acids Res* 2014, **42**(22):13674-13688.

197. Mosley AL HG, Sardiu ME, Smolle M, Workman JL, Florens L, Washburn MP.: **Quantitative Proteomics Demonstrates that the RNA Polymerase II Subunits Rpb4 and Rpb7 Dissociate During Transcription Elongation**. *Mol Cell Proteomics* 2013, **12**(6).

198. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H *et al*: **Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis**. *Science* 1999, **285**(5429):901-906.

199. Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B: **The tandem affinity purification (TAP) method: a general procedure of protein complex purification**. *Methods* 2001, **24**(3):218-229.

200. Funakoshi M, Hochstrasser M: **Small epitope-linker modules for PCR-based C-terminal tagging in Saccharomyces cerevisiae**. *Yeast* 2009, **26**(3):185-192.

201. Mosley AL, Florens L, Wen Z, Washburn MP: **A label free quantitative proteomic analysis of the Saccharomyces cerevisiae nucleus**. *J Proteomics* 2009, **72**(1):110-120.

202. Mosley ALS, Mihaela E.; Pattenden, Samantha G.; Workman, Jerry L.; Florens, Laurence; Washburn, Michael P.: **Highly reproducible label free quantitative proteomic analysis of RNA polymerase coplexes**. *Molecular and cellular proteomics* 2011, **10**(2).

203. Homer NM, Barry; Nelson, Stanley F.: **BFAST: an alignment tool for large scale genome resequencing**. *PLoS One* 2009, **4**(11):e7767.

204. Breese MR, Liu Y: **NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets**. *Bioinformatics* 2013, **29**(4):494-496.

205. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, **26**(1):139-140.

206. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer**. *Nat Biotechnol* 2011, **29**(1):24-26.

207. Thorvaldsdottir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration**. *Brief Bioinform* 2013, **14**(2):178-192.

208. Jenks MH, O'Rourke TW, Reines D: **Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast**. *Mol Cell Biol* 2008, **28**(12):3883-3893.

209. Beissbarth T, Speed TP: **GOstat: find statistically overrepresented Gene Ontologies within a group of genes**. *Bioinformatics* 2004, **20**(9):1464-1465.

210. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources**. *Nat Protoc* 2009, **4**(1):44-57.

211. Rhee HS, Pugh BF: **ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy**. *Curr Protoc Mol Biol* 2012, **Chapter 21**:Unit 21 24.

212. Trinh QM, Jen FY, Zhou Z, Chu KM, Perry MD, Kephart ET, Contrino S, Ruzanov P, Stein LD: **Cloud-based uniform ChIP-Seq processing tools for modENCODE and ENCODE**. *BMC Genomics* 2013, **14**:494.

213. Feng J, Liu T, Qin B, Zhang Y, Liu XS: **Identifying ChIP-seq enrichment using MACS**. *Nat Protoc* 2012, **7**(9):1728-1740.

214. Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J, Cramer P: **Transcriptome surveillance by selective termination of noncoding RNA synthesis**. *Cell* 2013, **155**(5):1075-1087.

215. Park D, Morris AR, Battenhouse A, Iyer VR: **Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements**. *Nucleic Acids Res* 2014.

216. van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Ne P, Loeillet S *et al*: **XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast**. *Nature* 2011, **475**(7354):114-117.

217. Castelnuovo M, Rahman S, Guffanti E, Infantino V, Stutz F, Zenklusen D: **Bimodal expression of PHO84 is modulated by early termination of antisense transcription**. *Nat Struct Mol Biol* 2013, **20**(7):851-858.

218. Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM, Proudfoot NJ: **Gene loops enhance transcriptional directionality**. *Science* 2012, **338**(6107):671-675.

219. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M: **The transcriptional landscape of the yeast genome defined by RNA sequencing**. *Science* 2008, **320**(5881):1344-1349.

220. Ganem C, Devaux F, Torchet C, Jacq C, Quevillon-Cheruel S, Labesse G, Facca C, Faye G: **Ssu72 is a phosphatase essential for transcription termination of snoRNAs and specific mRNAs in yeast**. *Embo J* 2003, **22**(7):1588-1598.

221. Marquardt S, Hazelbaker DZ, Buratowski S: **Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species**. *Transcription* 2011, **2**(3):145-154.

222. Loya TJ, O'Rourke TW, Degtyareva N, Reines D: **A network of interdependent molecular interactions describes a higher order Nrd1-Nab3 complex involved in yeast transcription termination**. *J Biol Chem* 2013.

223. Loya TJ, O'Rourke TW, Reines D: **A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3**. *Nucleic Acids Res* 2012.

224. Lowe TM, Eddy SR: **A computational screen for methylation guide snoRNAs in yeast**. *Science* 1999, **283**(5405):1168-1171.

225. Samarsky DA, Fournier MJ: **A comprehensive database for the small nucleolar RNAs from Saccharomyces cerevisiae**. *Nucleic Acids Res* 1999, **27**(1):161-164.

226. Camblong J, Beyrouthy N, Guffanti E, Schlaepfer G, Steinmetz LM, Stutz F: **Trans-acting antisense RNAs mediate transcriptional gene cosuppression in S. cerevisiae**. *Genes Dev* 2009, **23**(13):1534-1545.

227. Houseley J, Tollervey D: **The nuclear RNA surveillance machinery: the link between ncRNAs and genome structure in budding yeast?** *Biochim Biophys Acta* 2008, **1779**(4):239-246.

228. Uhler JP, Hertel C, Svejstrup JQ: **A role for noncoding transcription in activation of the yeast PHO5 gene**. *Proc Natl Acad Sci U S A* 2007, **104**(19):8011-8016.

229. Pelechano V, Steinmetz LM: **Gene regulation by antisense transcription**. *Nat Rev Genet* 2013, **14**(12):880-893.

230. Cloutier SC, Wang S, Ma WK, Petell CJ, Tran EJ: **Long noncoding RNAs promote transcriptional poising of inducible genes**. *PLoS Biol* 2013, **11**(11):e1001715.

231. Terzi N, Churchman LS, Vasiljeva L, Weissman J, Buratowski S: **H3K4 trimethylation by Set1 promotes efficient termination by the Nrd1-Nab3-Sen1 pathway**. *Mol Cell Biol* 2011, **31**(17):3569-3583.

232. Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnirke A, Nusbaum C, Thompson DA, Friedman N, Regev A: **Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species**. *Genome Biol* 2010, **11**(8):R87.

233. Lenstra TL, Tudek A, Clauder S, Xu Z, Pachis ST, van Leenen D, Kemmeren P, Steinmetz LM, Libri D, Holstege FC: **The Role of Ctk1**

Kinase in Termination of Small Non-Coding RNAs. *PLoS One* 2013, **8**(12):e80495.

234. Nadkarni AK, McDonough VM, Yang WL, Stukey JE, Ozier-Kalogeropoulos O, Carman GM: **Differential biochemical regulation of the URA7- and URA8-encoded CTP synthetases from Saccharomyces cerevisiae**. *J Biol Chem* 1995, **270**(42):24982-24988.

235. Venema J, Tollervey D: **Ribosome synthesis in Saccharomyces cerevisiae**. *Annu Rev Genet* 1999, **33**:261-311.

236. Kwapisz M, Wery M, Despres D, Ghavi-Helm Y, Soutourina J, Thuriaux P, Lacroute F: **Mutations of RNA polymerase II activate key genes of the nucleoside triphosphate biosynthetic pathways**. *Embo J* 2008, **27**(18):2411-2421.

237. Sardiu MEF, Laurence; Washburn, Michael P.: **Evaluation of clustering algorithms for protein complex and protein interaction network assembly**. *JOurnal of proteome research* 2009, **8**:2944-2952.

238. Nordick K, Hoffman MG, Betz JL, Jaehning JA: **Direct interactions between the Paf1 complex and a cleavage and polyadenylation factor are revealed by dissociation of Paf1 from RNA polymerase II**. *Eukaryot Cell* 2008, **7**(7):1158-1167.

239. Mueller CL, Porter SE, Hoffman MG, Jaehning JA: **The Paf1 complex has functions independent of actively transcribing RNA polymerase II**. *Mol Cell* 2004, **14**(4):447-456.

240. Penheiter KL, Washburn TM, Porter SE, Hoffman MG, Jaehning JA: **A posttranscriptional role for the yeast Paf1-RNA polymerase II complex is revealed by identification of primary targets**. *Mol Cell* 2005, **20**(2):213-223.

241. Sheldon KE, Mauger DM, Arndt KM: **A Requirement for the Saccharomyces cerevisiae Paf1 complex in snoRNA 3' end formation**. *Mol Cell* 2005, **20**(2):225-236.

242. Tomson BN, Davis CP, Warner MH, Arndt KM: **Identification of a role for histone H2B ubiquitylation in noncoding RNA 3'-end formation through mutational analysis of Rtf1 in Saccharomyces cerevisiae**. *Genetics* 2011, **188**(2):273-289.

243. Fox MJ, Gao H, Smith-Kinnaman WR, Liu Y, Mosley AL: **The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway**. *PLoS Genet* 2015, **10**(2):e1004999.

244. Shaw RJ, Wilson JL, Smith KT, Reines D: **Regulation of an IMP dehydrogenase gene and its overexpression in drug-sensitive transcription elongation mutants of yeast**. *J Biol Chem* 2001, **276**(35):32905-32916.

245. Sloan KE, Schneider C, Watkins NJ: **Comparison of the yeast and human nuclear exosome complexes**. *Biochem Soc Trans* 2012, **40**(4):850-855.

246. Fasken MB, Leung SW, Banerjee A, Kodani MO, Chavez R, Bowman EA, Purohit MK, Rubinson ME, Rubinson EH, Corbett AH: **Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex**. *J Biol Chem* 2011, **286**(43):37429-37445.

247. Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falkenby LG, Lykke-Andersen S, Andersen JS, Dziembowski A, Jensen TH: **Interaction profiling identifies the human nuclear exosome targeting complex**. *Mol Cell* 2011, **43**(4):624-637.

248. Hrossova D, Sikorsky T, Potesil D, Bartosovic M, Pasulka J, Zdrahal Z, Stefl R, Vanacova S: **RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs**. *Nucleic Acids Res* 2015.

249. Gornemann J, Kotovic KM, Hujer K, Neugebauer KM: **Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex**. *Mol Cell* 2005, **19**(1):53-63.

250. Flaherty SM, Fortes P, Izaurralde E, Mattaj IW, Gilmartin GM: **Participation of the nuclear cap binding complex in pre-mRNA 3' processing**. *Proc Natl Acad Sci U S A* 1997, **94**(22):11893-11898.

251. Hosoda N, Kim YK, Lejeune F, Maquat LE: **CBP80 promotes interaction of Upf1 with Upf2 during nonsense-mediated mRNA decay in mammalian cells**. *Nat Struct Mol Biol* 2005, **12**(10):893-901.

252. Andersen PR, Domanski M, Kristiansen MS, Storvall H, Ntini E, Verheggen C, Schein A, Bunkenborg J, Poser I, Hallais M *et al*: **The human cap-binding complex is functionally connected to the nuclear RNA exosome**. *Nat Struct Mol Biol* 2013, **20**(12):1367-1376.

253. Hallais M, Pontvianne F, Andersen PR, Clerici M, Lener D, Benbahouche Nel H, Gostan T, Vandermoere F, Robert MC, Cusack S *et al*: **CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing**. *Nat Struct Mol Biol* 2013, **20**(12):1358-1366.

254. Kuai L, Das B, Sherman F: **A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae**. *Proc Natl Acad Sci U S A* 2005, **102**(39):13962-13967.

255. Das B, Butler JS, Sherman F: **Degradation of normal mRNA in the nucleus of Saccharomyces cerevisiae**. *Mol Cell Biol* 2003, **23**(16):5502-5515.

256. Tian B, Manley JL: **Alternative cleavage and polyadenylation: the long and short of it**. *Trends Biochem Sci* 2013, **38**(6):312-320.

257. Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL *et al*: **Systematic profiling of poly(A)+ transcripts modulated by core 3' end processing and splicing factors reveals regulatory rules of alternative cleavage and polyadenylation**. *PLoS Genet* 2015, **11**(4):e1005166.

258. Akman HB, Erson-Bensan AE: **Alternative polyadenylation and its impact on cellular processes**. *Microrna* 2014, **3**(1):2-9.

259. Curinha A, Oliveira Braz S, Pereira-Castro I, Cruz A, Moreira A: **Implications of polyadenylation in health and disease**. *Nucleus* 2014, **5**(6):508-519.

260. Berkovits BD, Mayr C: **Alternative 3' UTRs act as scaffolds to regulate membrane protein localization**. *Nature* 2015, **522**(7556):363-367.

261. Ni T, Yang Y, Hafez D, Yang W, Kiesewetter K, Wakabayashi Y, Ohler U, Peng W, Zhu J: **Distinct polyadenylation landscapes of diverse human tissues revealed by a modified PA-seq strategy**. *BMC Genomics* 2013, **14**:615.

262. Blechingberg J, Lykke-Andersen S, Jensen TH, Jorgensen AL, Nielsen AL: **Regulatory mechanisms for 3'-end alternative splicing and polyadenylation of the Glial Fibrillary Acidic Protein, GFAP, transcript**. *Nucleic Acids Res* 2007, **35**(22):7636-7650.

263. Brown SL, Morrison SL: **Developmental regulation of membrane and secretory Ig gamma 2b mRNA**. *J Immunol* 1989, **142**(6):2072-2080.

264. Peterson ML: **Mechanisms controlling production of membrane and secreted immunoglobulin during B cell development**. *Immunol Res* 2007, **37**(1):33-46.

265. Peterson ML: **Immunoglobulin heavy chain gene regulation through polyadenylation and splicing competition**. *Wiley Interdiscip Rev RNA* 2011, **2**(1):92-105.

266. Dass B, Tardif S, Park JY, Tian B, Weitlauf HM, Hess RA, Carnes K, Griswold MD, Small CL, Macdonald CC: **Loss of polyadenylation**

protein tauCstF-64 causes spermatogenic defects and male infertility. *Proc Natl Acad Sci U S A* 2007, **104**(51):20374-20379.

267. Hockert KJ, Martincic K, Mendis-Handagama SM, Borghesi LA, Milcarek C, Dass B, MacDonald CC: **Spermatogenetic but not immunological defects in mice lacking the tauCstF-64 polyadenylation protein**. *J Reprod Immunol* 2011, **89**(1):26-37.

268. Li W, Yeh HJ, Shankarling GS, Ji Z, Tian B, MacDonald CC: **The tauCstF-64 polyadenylation protein controls genome expression in testis**. *PLoS One* 2012, **7**(10):e48373.

269. Liu D, Brockman JM, Dass B, Hutchins LN, Singh P, McCarrey JR, MacDonald CC, Graber JH: **Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis**. *Nucleic Acids Res* 2007, **35**(1):234-246.

270. Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC: **Widespread and extensive lengthening of 3' UTRs in the mammalian brain**. *Genome Res* 2013, **23**(5):812-825.

271. Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B: **Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing**. *Nat Methods* 2013, **10**(2):133-139.

272. Shepard PJ, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y: **Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq**. *Rna* 2011, **17**(4):761-772.

273. Hinske LC, Galante PA, Limbeck E, Mohnle P, Parmigiani RB, Ohno-Machado L, Camargo AA, Kreth S: **Alternative polyadenylation allows differential negative feedback of human miRNA miR-579 on its host gene ZFR**. *PLoS One* 2015, **10**(3):e0121507.

274. Lai DP, Tan S, Kang YN, Wu J, Ooi HS, Chen J, Shen TT, Qi Y, Zhang X, Guo Y *et al*: **Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis**. *Hum Mol Genet* 2015, **24**(12):3410-3417.

275. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB: **Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites**. *Science* 2008, **320**(5883):1643-1647.

276. Elkon R, Drost J, van Haaften G, Jenal M, Schrier M, Oude Vrielink JA, Agami R: **E2F mediates enhanced alternative polyadenylation in proliferation**. *Genome Biol* 2012, **13**(7):R59.

277. Vindry C, Vo Ngoc L, Kruys V, Gueydan C: **RNA-binding protein-mediated post-transcriptional controls of gene expression: integration of molecular mechanisms at the 3' end of mRNAs?** *Biochem Pharmacol* 2014, **89**(4):431-440.

278. Hollerer I, Grund K, Hentze MW, Kulozik AE: **mRNA 3'end processing: A tale of the tail reaches the clinic**. *EMBO Mol Med* 2014, **6**(1):16-26.

279. Boutet SC, Cheung TH, Quach NL, Liu L, Prescott SL, Edalati A, Iori K, Rando TA: **Alternative polyadenylation mediates microRNA**

regulation of muscle stem cell function. *Cell Stem Cell* 2012, **10**(3):327-336.

280. Natalizio BJ, Muniz LC, Arhin GK, Wilusz J, Lutz CS: **Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals**. *J Biol Chem* 2002, **277**(45):42733-42740.

281. Hall-Pogar T, Liang S, Hague LK, Lutz CS: **Specific trans-acting proteins interact with auxiliary RNA polyadenylation elements in the COX-2 3'-UTR**. *Rna* 2007, **13**(7):1103-1115.

282. Brackenridge S, Proudfoot NJ: **Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal**. *Mol Cell Biol* 2000, **20**(8):2660-2669.

283. Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ: **The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms**. *Genes Dev* 1998, **12**(16):2522-2534.

284. Danckwardt S, Kaufmann I, Gentzel M, Foerstner KU, Gantzert AS, Gehring NH, Neu-Yilik G, Bork P, Keller W, Wilm M *et al*: **Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals**. *Embo J* 2007, **26**(11):2658-2669.

285. Danckwardt S, Gantzert AS, Macher-Goeppinger S, Probst HC, Gentzel M, Wilm M, Grone HJ, Schirmacher P, Hentze MW, Kulozik AE: **p38**

MAPK controls prothrombin expression by regulated RNA 3' end processing. *Mol Cell* 2011, **41**(3):298-310.

286. Danckwardt S, Gehring NH, Neu-Yilik G, Hundsdoerfer P, Pforsich M, Frede U, Hentze MW, Kulozik AE: **The prothrombin 3'end formation signal reveals a unique architecture that is sensitive to thrombophilic gain-of-function mutations**. *Blood* 2004, **104**(2):428-435.

287. Gehring NH, Frede U, Neu-Yilik G, Hundsdoerfer P, Vetter B, Hentze MW, Kulozik AE: **Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia**. *Nat Genet* 2001, **28**(4):389-392.

288. Stacey SN, Sulem P, Jonasdottir A, Masson G, Gudmundsson J, Gudbjartsson DF, Magnusson OT, Gudjonsson SA, Sigurgeirsson B, Thorisdottir K *et al*: **A germline variant in the TP53 polyadenylation signal confers cancer susceptibility**. *Nat Genet* 2011, **43**(11):1098-1103.

289. Enciso-Mora V, Hosking FJ, Di Stefano AL, Zelenika D, Shete S, Broderick P, Idbaih A, Delattre JY, Hoang-Xuan K, Marie Y *et al*: **Low penetrance susceptibility to glioma is caused by the TP53 variant rs78378222**. *Br J Cancer* 2013, **108**(10):2178-2185.

290. Bennett CL, Ochs HD: **IPEX is a unique X-linked syndrome characterized by immune dysfunction, polyendocrinopathy, enteropathy, and a variety of autoimmune phenomena**. *Curr Opin Pediatr* 2001, **13**(6):533-538.

291. Weinstein IB: **Cancer. Addiction to oncogenes--the Achilles heal of cancer**. *Science* 2002, **297**(5578):63-64.

292. Chao SH, Fujinaga K, Marion JE, Taube R, Sausville EA, Senderowicz AM, Peterlin BM, Price DH: **Flavopiridol inhibits P-TEFb and blocks HIV-1 replication**. *J Biol Chem* 2000, **275**(37):28345-28348.

293. Chen R, Gandhi V, Plunkett W: **A sequential blockade strategy for the design of combination therapies to overcome oncogene addiction in chronic myelogenous leukemia**. *Cancer Res* 2006, **66**(22):10959-10966.

294. Wasylishen AR, Penn LZ: **Myc: the beauty and the beast**. *Genes Cancer* 2010, **1**(6):532-541.

295. Eilers M, Eisenman RN: **Myc's broad reach**. *Genes Dev* 2008, **22**(20):2755-2766.

296. Nie Z, Hu G, Wei G, Cui K, Yamane A, Resch W, Wang R, Green DR, Tessarollo L, Casellas R *et al*: **c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells**. *Cell* 2012, **151**(1):68-79.

297. Schuhmacher M, Eick D: **Dose-dependent regulation of target gene expression and cell proliferation by c-Myc levels**. *Transcription* 2013, **4**(4):192-197.

298. Loven J, Orlando DA, Sigova AA, Lin CY, Rahl PB, Burge CB, Levens DL, Lee TI, Young RA: **Revisiting global gene expression analysis**. *Cell* 2012, **151**(3):476-482.

299. Lin CY, Loven J, Rahl PB, Paranal RM, Burge CB, Bradner JE, Lee TI, Young RA: **Transcriptional amplification in tumor cells with elevated c-Myc**. *Cell* 2012, **151**(1):56-67.

300. Chipumuro E, Marco E, Christensen CL, Kwiatkowski N, Zhang T, Hatheway CM, Abraham BJ, Sharma B, Yeung C, Altabef A *et al*: **CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer**. *Cell* 2014, **159**(5):1126-1139.

301. Glover-Cutter K, Larochelle S, Erickson B, Zhang C, Shokat K, Fisher RP, Bentley DL: **TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II**. *Mol Cell Biol* 2009, **29**(20):5455-5464.

302. Chen X, Muller U, Sundling KE, Brow DA: **Saccharomyces cerevisiae Sen1 as a model for the study of mutations in human Senataxin that elicit cerebellar ataxia**. *Genetics* 2014, **198**(2):577-590.

303. Bennett CL, La Spada AR: **Unwinding the role of senataxin in neurodegeneration**. *Discov Med* 2015, **19**(103):127-136.

304. Lemmens R, Moore MJ, Al-Chalabi A, Brown RH, Jr., Robberecht W: **RNA metabolism and the pathogenesis of motor neuron diseases**. *Trends Neurosci* 2010, **33**(5):249-258.

305. Chen YZ, Bennett CL, Huynh HM, Blair IP, Puls I, Irobi J, Dierick I, Abel A, Kennerson ML, Rabin BA *et al*: **DNA/RNA helicase gene mutations in a**

form of juvenile amyotrophic lateral sclerosis (ALS4). *Am J Hum Genet* 2004, **74**(6):1128-1135.

306. Moreira MC, Klur S, Watanabe M, Nemeth AH, Le Ber I, Moniz JC, Tranchant C, Aubourg P, Tazir M, Schols L *et al*: **Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2**. *Nat Genet* 2004, **36**(3):225-227.

307. Mischo HE, Gomez-Gonzalez B, Grzechnik P, Rondon AG, Wei W, Steinmetz L, Aguilera A, Proudfoot NJ: **Yeast Sen1 helicase protects the genome from transcription-associated instability**. *Mol Cell* 2011, **41**(1):21-32.

308. Alzu A, Bermejo R, Begnis M, Lucca C, Piccini D, Carotenuto W, Saponaro M, Brambati A, Cocito A, Foiani M *et al*: **Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes**. *Cell* 2012, **151**(4):835-846.

309. Golla U, Singh V, Azad GK, Singh P, Verma N, Mandal P, Chauhan S, Tomar RS: **Sen1p contributes to genomic integrity by regulating expression of ribonucleotide reductase 1 (RNR1) in Saccharomyces cerevisiae**. *PLoS One* 2013, **8**(5):e64798.

310. Chavez S, Beilharz T, Rondon AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A: **A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae**. *Embo J* 2000, **19**(21):5824-5834.

311. Gallardo M, Aguilera A: **A new hyperrecombination mutation identifies a novel yeast gene, THP1, connecting transcription elongation with mitotic recombination**. *Genetics* 2001, **157**(1):79-89.

312. Fischer T, Strasser K, Racz A, Rodriguez-Navarro S, Oppizzi M, Ihrig P, Lechner J, Hurt E: **The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores**. *Embo J* 2002, **21**(21):5843-5852.

313. Huertas P, Aguilera A: **Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination**. *Mol Cell* 2003, **12**(3):711-721.

314. Li X, Manley JL: **Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability**. *Cell* 2005, **122**(3):365-378.

315. Gonzalez-Aguilera C, Tous C, Gomez-Gonzalez B, Huertas P, Luna R, Aguilera A: **The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability**. *Mol Biol Cell* 2008, **19**(10):4310-4318.

316. Libri D, Dower K, Boulay J, Thomsen R, Rosbash M, Jensen TH: **Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation**. *Mol Cell Biol* 2002, **22**(23):8254-8266.

317. Rondon AG, Jimeno S, Garcia-Rubio M, Aguilera A: **Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation**. *J Biol Chem* 2003, **278**(40):39037-39043.

318. Mason PB, Struhl K: **Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo**. *Mol Cell* 2005, **17**(6):831-840.

319. Wellinger RE, Prado F, Aguilera A: **Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex**. *Mol Cell Biol* 2006, **26**(8):3327-3334.

320. Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C, Boulay J, Jensen TH, Stutz F, Devaux F *et al*: **THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association**. *Cell* 2008, **135**(2):308-321.

321. Yuce O, West SC: **Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response**. *Mol Cell Biol* 2013, **33**(2):406-417.

322. Becherel OJ, Yeo AJ, Stellati A, Heng EY, Luff J, Suraweera AM, Woods R, Fleming J, Carrie D, McKinney K *et al*: **Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing**. *PLoS Genet* 2013, **9**(4):e1003435.

323. Franchitto A: **Genome instability at common fragile sites: searching for the cause of their instability**. *Biomed Res Int* 2013, **2013**:730714.

324. Durkin SG, Glover TW: **Chromosome fragile sites**. *Annu Rev Genet* 2007, **41**:169-192.

325. Richard P, Manley JL: **SETX sumoylation: A link between DNA damage and RNA surveillance disrupted in AOA2**. *Rare Dis* 2014, **2**:e27744.

326. Gelpi C, Alguero A, Angeles Martinez M, Vidal S, Juarez C, Rodriguez-Sanchez JL: **Identification of protein components reactive with anti-PM/Scl autoantibodies**. *Clin Exp Immunol* 1990, **81**(1):59-64.

327. Reichlin M, Maddison PJ, Targoff I, Bunch T, Arnett F, Sharp G, Treadwell E, Tan EM: **Antibodies to a nuclear/nucleolar antigen in patients with polymyositis overlap syndromes**. *J Clin Immunol* 1984, **4**(1):40-44.

328. Mahler M, Raijmakers R: **Novel aspects of autoantibodies to the PM/Scl complex: clinical, genetic and diagnostic insights**. *Autoimmun Rev* 2007, **6**(7):432-437.

329. Airoldi G, Guidarelli A, Cantoni O, Panzeri C, Vantaggiato C, Bonato S, Grazia D'Angelo M, Falcone S, De Palma C, Tonelli A *et al*: **Characterization of two novel SETX mutations in AOA2 patients reveals aspects of the pathophysiological role of senataxin**. *Neurogenetics* 2010, **11**(1):91-100.

330. Suraweera A, Lim Y, Woods R, Birrell GW, Nasim T, Becherel OJ, Lavin MF: **Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation**. *Hum Mol Genet* 2009, **18**(18):3384-3396.

331. Roberts TC, Morris KV, Wood MJ: **The role of long non-coding RNAs in neurodevelopment, brain function and neurological disease**. *Philos Trans R Soc Lond B Biol Sci* 2014, **369**(1652).

332. van de V, II, Gordebeke PM, Khoshab N, Tiesinga PH, Buitelaar JK, Kozicz T, Aschrafi A, Glennon JC: **Long non-coding RNAs in neurodevelopmental disorders**. *Front Mol Neurosci* 2013, **6**:53.

333. Shi X, Sun M, Liu H, Yao Y, Song Y: **Long non-coding RNAs: a new frontier in the study of human diseases**. *Cancer Lett* 2013, **339**(2):159-166.

334. Verdel A, Moazed D: **RNAi-directed assembly of heterochromatin in fission yeast**. *FEBS Lett* 2005, **579**(26):5872-5878.

335. Lum PY, Armour CD, Stepaniants SB, Cavet G, Wolf MK, Butler JS, Hinshaw JC, Garnier P, Prestwich GD, Leonardson A *et al*: **Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes**. *Cell* 2004, **116**(1):121-137.

336. Marin-Vicente C, Domingo-Prim J, Eberle AB, Visa N: **RRP6/EXOSC10 is required for the repair of DNA double-strand breaks by homologous recombination**. *J Cell Sci* 2015, **128**(6):1097-1107.

337. Andrulis ED, Werner J, Nazarian A, Erdjument-Bromage H, Tempst P, Lis JT: **The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila**. *Nature* 2002, **420**(6917):837-841.

338. Lemay JF, Larochelle M, Marguerat S, Atkinson S, Bahler J, Bachand F: **The RNA exosome promotes transcription termination of backtracked RNA polymerase II**. *Nat Struct Mol Biol* 2014, **21**(10):919-926.

339. Gomez-Herreros F, de Miguel-Jimenez L, Millan-Zambrano G, Penate X, Delgado-Ramos L, Munoz-Centeno MC, Chavez S: **One step back before moving forward: regulation of transcription elongation by arrest and backtracking**. *FEBS Lett* 2012, **586**(18):2820-2825.

340. Cheung AC, Cramer P: **Structural basis of RNA polymerase II backtracking, arrest and reactivation**. *Nature* 2011, **471**(7337):249-253.

341. Eckmann CR, Rammelt C, Wahle E: **Control of poly(A) tail length**. *Wiley Interdiscip Rev RNA* 2011, **2**(3):348-361.

342. Amrani N, Minet M, Le Gouar M, Lacroute F, Wyers F: **Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro**. *Mol Cell Biol* 1997, **17**(7):3694-3701.

343. Brune C, Munchel SE, Fischer N, Podtelejnikov AV, Weis K: **Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export**. *Rna* 2005, **11**(4):517-531.

344. Minvielle-Sebastia L, Preker PJ, Wiederkehr T, Strahm Y, Keller W: **The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation**. *Proc Natl Acad Sci U S A* 1997, **94**(15):7897-7902.

345. Fasken MB, Stewart M, Corbett AH: **Functional significance of the interaction between the mRNA-binding protein, Nab2, and the nuclear pore-associated protein, Mlp1, in mRNA export**. *J Biol Chem* 2008, **283**(40):27130-27143.

346. Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR, Wilson SM, Minvielle-Sebastia L, Swanson MS: **Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export**. *Embo J* 2002, **21**(7):1800-1810.

347. Iglesias N, Tutucci E, Gwizdek C, Vinciguerra P, Von Dach E, Corbett AH, Dargemont C, Stutz F: **Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export**. *Genes Dev* 2010, **24**(17):1927-1938.

348. Lemay JF, Lemieux C, St-Andre O, Bachand F: **Crossing the borders: poly(A)-binding proteins working on both sides of the fence**. *RNA Biol* 2010, **7**(3):291-295.

349. Roth KM, Wolf MK, Rossi M, Butler JS: **The nuclear exosome contributes to autogenous control of NAB2 mRNA levels**. *Mol Cell Biol* 2005, **25**(5):1577-1585.

350. Dez C, Houseley J, Tollervey D: **Surveillance of nuclear-restricted pre-ribosomes within a subnucleolar region of Saccharomyces cerevisiae**. *Embo J* 2006, **25**(7):1534-1546.

351. Wery M, Ruidant S, Schillewaert S, Lepore N, Lafontaine DL: **The nuclear poly(A) polymerase and Exosome cofactor Trf5 is recruited cotranscriptionally to nucleolar surveillance**. *Rna* 2009, **15**(3):406-419.

352. Legrain P, Rosbash M: **Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm**. *Cell* 1989, **57**(4):573-583.

353. Shiomi T, Fukushima K, Suzuki N, Nakashima N, Noguchi E, Nishimoto T: **Human dis3p, which binds to either GTP- or GDP-Ran, complements Saccharomyces cerevisiae dis3**. *J Biochem* 1998, **123**(5):883-890.

354. Bousquet-Antonelli C, Presutti C, Tollervey D: **Identification of a regulated pathway for nuclear pre-mRNA turnover**. *Cell* 2000, **102**(6):765-775.

355. Hilleren P, McCarthy T, Rosbash M, Parker R, Jensen TH: **Quality control of mRNA 3'-end processing is linked to the nuclear exosome**. *Nature* 2001, **413**(6855):538-542.

356. Hilleren P, Parker R: **Defects in the mRNA export factors Rat7p, Gle1p, Mex67p, and Rat8p cause hyperadenylation during 3'-end formation of nascent transcripts**. *Rna* 2001, **7**(5):753-764.

357. Jensen TH, Patricio K, McCarthy T, Rosbash M: **A block to mRNA nuclear export in S. cerevisiae leads to hyperadenylation of transcripts that accumulate at the site of transcription**. *Mol Cell* 2001, **7**(4):887-898.

358. Torchet C, Bousquet-Antonelli C, Milligan L, Thompson E, Kufel J, Tollervey D: **Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs**. *Mol Cell* 2002, **9**(6):1285-1296.

359. Synowsky SA, Heck AJ: **The yeast Ski complex is a hetero-tetramer**. *Protein Sci* 2008, **17**(1):119-125.

360. Anderson JS, Parker RP: **The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box protein and 3' to 5' exonucleases of the exosome complex**. *Embo J* 1998, **17**(5):1497-1506.

361. Houseley J, Tollervey D: **The many pathways of RNA degradation**. *Cell* 2009, **136**(4):763-776.

362. Parker R: **RNA degradation in Saccharomyces cerevisae**. *Genetics* 2012, **191**(3):671-702.

363. Schmid M, Jensen TH: **The exosome: a multipurpose RNA-decay machine**. *Trends Biochem Sci* 2008, **33**(10):501-510.

364. van Hoof A, Frischmeyer PA, Dietz HC, Parker R: **Exosome-mediated recognition and degradation of mRNAs lacking a termination codon**. *Science* 2002, **295**(5563):2262-2264.

365. Schaeffer D, Clark A, Klauer AA, Tsanova B, van Hoof A: **Functions of the cytoplasmic exosome**. *Adv Exp Med Biol* 2011, **702**:79-90.

366. Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC: **An mRNA surveillance mechanism that eliminates transcripts lacking termination codons**. *Science* 2002, **295**(5563):2258-2261.

367. Warner JR: **The economics of ribosome biosynthesis in yeast**. *Trends Biochem Sci* 1999, **24**(11):437-440.

368. Harigaya Y, Tanaka H, Yamanaka S, Tanaka K, Watanabe Y, Tsutsumi C, Chikashige Y, Hiraoka Y, Yamashita A, Yamamoto M: **Selective**

elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* 2006, **442**(7098):45-50.

369. McIver SC, Kang YA, DeVilbiss AW, O'Driscoll CA, Ouellette JN, Pope NJ, Camprecios G, Chang CJ, Yang D, Bouhassira EE *et al*: **The exosome complex establishes a barricade to erythroid maturation**. *Blood* 2014, **124**(14):2285-2297.

370. Yamazaki T, Hirose T: **The building process of the functional paraspeckle with long non-coding RNAs**. *Front Biosci (Elite Ed)* 2015, **7**:1-41.

371. Singh BN, Hampsey M: **A transcription-independent role for TFIIB in gene looping**. *Mol Cell* 2007, **27**(5):806-816.

372. Ansari A, Hampsey M: **A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping**. *Genes Dev* 2005, **19**(24):2969-2978.

373. Krishnamurthy S, Ghazy MA, Moore C, Hampsey M: **Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries**. *Mol Cell Biol* 2009, **29**(11):2925-2934.

374. Werner-Allen JW, Lee CJ, Liu P, Nicely NI, Wang S, Greenleaf AL, Zhou P: **cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72**. *J Biol Chem* 2011, **286**(7):5717-5726.

375. Chinchilla K, Rodriguez-Molina JB, Ursic D, Finkel JS, Ansari AZ, Culbertson MR: **Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae**. *Eukaryot Cell* 2012, **11**(4):417-429.

CURRICULUM VITAE

Melanie J. Fox

EDUCATION

Ph.D., Biochemistry & Molecular Biology (2015)
Indiana University, Indianapolis, Indiana

Bachelor of Arts, Biology (2008)
Bluffton University, Bluffton Ohio

RESEARCH EXPERIENCE

Ph.D Graduate Student, Biochemistry and Molecular Biology
Indiana University School of Medicine, Indianapolis, Indiana
August 2010 – 2015

Adviser: Dr. Amber Mosley
Dissertation: The Role of Rtr1 and Rrp6 in RNAPII Transcription Termination
Investigated the role of the RNA Polymerase II C-Terminal Domain (CTD) Phosphatase Rtr1 and the nuclear RNA exonuclease Rrp6 in transcription termination of short RNAs using genome- and transcriptome-wide sequencing

Research Technician & Laboratory Manager, Medical & Molecular Genetics
Indiana University School of Medicine, Indianapolis, Indiana
June 2008 – August 2010

Advisor: Dr. Brittany-Shea Herbert
Investigated the effect of small molecular inhibitors and miRNAs targeting telomerase in breast and ovarian cancer cell lines, breast cancer mouse models, and circulating tumor cells in a clinical trial

Undergraduate Research, Independent Study Program
Bluffton University, Bluffton, Ohio
August 2007 – May 2008

Advisor: Dr. Stephen Steiner
Investigated the role of Bowman-Birk protease inhibitors in legumes using basic techniques such as DNA isolation, Polymerase Chain Reaction (PCR), and molecular cloning

Techniques Utilized

Mammalian cell culture – culture maintenance, cell proliferation, colony formation, RNA isolation
Yeast culture – genetic manipulation, DNA and RNA isolation, TAP- and FLAG-purification

Small animal handling – subcutaneous *in vivo* tumor models, subcutaneous and intraperitoneal injections
Chromatin immunoprecipitation (ChIP), ChIP-Exo followed by DNA-sequencing
Northern and Western Immunoblotting
Silverstaining of Protein Gels
Immunofluorescence
PCR, RNA-Sequencing
Software: Integrative Genomics Viewer (IGV), Adobe Photoshop, Adobe Illustrator, Microsoft Office

PEER REVIEWED PUBLICATIONS

1. **Fox MJ**, Mosley AL. Rrp6: Integrated Roles in Nuclear RNA and Transcription. *Wiley Interdisciplinary Reviews: RNA*, In preparation.
2. **Fox MJ**, Smith-Kinnaman WR, Hunter GO, Gao H, Liu Y, Mosley AL. Rtr1 Regulates RNA Polymerase II Transcription Termination Through the Nrd1-Nab3 Pathway. Manuscript in preparation.
3. **Fox MJ**, Gao H, Smith-Kinnaman WR, Liu Y, Mosley AL. Characterization of Extended Transcripts by RNA-Seq suggests that Rrp6 regulates Nrd1-dependent Termination. *PLoS Genetics*, 2015.
4. Sauder CAM, Koziel JE, Choi M, **Fox MJ**, Grimes BR, Badve S, Blosser RJ, Radovish M, Lam CC, Vaughan MB, Herbert B-S, Clare SE. Phenotypic Plasticity in Normal Breast Derived Epithelial Cells. *BMC Cell Biology*. 15(20), 2014.
5. Smith-Kinnaman WR, Berna MJ, Hunter GO, Hsu P, True JD, Cabello GI, **Fox MJ**, Varani G, Mosley AL. The Interactome of the Atypical Phosphatase Rtr1 in *Saccharomyces cerevisiae*. *Molecular Biosystems*. 10(7):1730-41, 2014.
6. Koziel JE, **Fox MJ**, Steding CE, Sprouse AA, Herbert BS. Medical Genetics and Epigenetics of Telomerase. *Journal of Cellular and Molecular Medicine*. 15(3):457-67, 2011.
7. Goldblatt E, Gentry E, **Fox MJ,** Gryaznov S, Changyu S, Herbert B-S. The Telomerase Template Antagonist GRN163L Alters MDA-MB-231 Breast Cancer Cell Morphology, Inhibits Growth, and Augments the Effects of Paclitaxel. *Molecular Cancer Therapeutics*: 2027-2035, 2009.

POSTERS AND PRESENTATIONS

1. **Fox MJ**, Gao H, Liu Y, Mosley AL. "RNA-Seq Analysis Reveals Transcriptome-wide Effects of *RRP6* Deletion in *Saccharomyces cerevisiae," EMBO Conference Series: Gene Transcription in Yeast*, Sant Feliu de Guixols, Spain, June 2014
2. **Fox MJ**, Gao H, Liu Y, Mosley AL. Transcriptome-wide analysis of RNA processing defects in *S. cerevisiae*. *Rustbelt RNA Meeting*, Cleveland OH, October 2013.

3. Smith-Kinnaman WR, **Fox MJ**, Zimmerly M, Mosley AL. RNA Polymerase II and 3' end processing machinery in *S. cerevisiae*. *Rustbelt RNA Meeting*, Cleveland, OH, October 2013.
4. **Fox MJ**, Smith-Kinnaman WR, Hunter GO, Gao H, True JD, Fleharty B, Gogol M, Liu Y, Mosley A. The phosphatase Rtr1 regulates histone modification levels at transcription terminators. *Walther Cancer Foundation Symposium*, South Bend, IN, October 2013.
5. **Fox MJ**, Zimmerly MA, Hunter GO, Berna MJ, Smith-Kinnaman WR, Mosley AL. The role of CTD phosphorylation on mRNA 3'end processing machinery. *IUSM IBMG Program for PhD study interview weekend poster session.* Indianapolis, IN, January 2012.
6. Cabello G, **Fox M**, Hunter G, True J, Mosley A. The Role of Rtr1 in the Regulation of Genomic Integrity. *Indianapolis Project Seed Summer Research Program Poster Session*, Indianapolis, IN, July 2011.
7. Marcum CA, Blosser RJ, **Fox MJ**, Badve S, Herbert BS, Clare SE. Hidden in Plain Sight: Phenotypic Plasticity in the Normal Breast and the Origin of Rare Malignant Histologies. *J.K. Berman Essay Program*. Indianapolis, IN, June 2010.
8. Herbert B-S**, Fox MJ**, Skillman S. Role of microRNAs in the Regulation of Telomerase (hTERT) in Ovarian Cancer. *2009 ASHG Annual Meeting,* Hawaii, October 2009.
9. Skillman S, **Fox MJ**, Herbert B-S. The Role of MicroRNAs in the Regulation of Telomerase (hTERT) in Ovarian Cancer. *Student Research Program in Academic Medicine,* Indianapolis, IN, October 2009.

PROFESSIONAL ORGANIZATIONS

2013 – Present *Member,* American Society of Biochemistry and Molecular Biology
2013 – Present *Student Member,* Association of Science and Technology Centers

AWARDS & HONORS

2015: Erica Daniel Kepner Award for Scientific Achievement, Honorable Mention, IU School of Medicine
2015 William M. Plater Civic Engagement Medallion, IUPUI Office of Community Engagement
2014: Honorable Mention, Poster Presentation, IU School of Medicine Department of Biochemistry & Molecular Biology Research Day
2014: IUPUI Graduate & Professional Student Government Educational Enhancement Grant
Awarded for travel to EMBO Transcription Conference, Sant Feliu de Guixols, Spain
2010: IU Graduate School University Fellowship Award, for outstanding incoming students

TEACHING EXPERIENCE

Guest Lecturer

(BIOC-848) Bioinformatic applications to proteomics and genomics
Indiana University School of Medicine
Indianapolis, Indiana

Course Director: Dr. Amber Mosley
Lecture Title: RNA-Sequencing
Date: March 4, 2015

Student Mentor

Herbert and Mosley Labs, Indiana University School of Medicine
Indianapolis, Indiana
September 2008 – January 2015

- Supervised four laboratory interns, including two high school students and two undergraduate students
- Mentored students through independent research projects and final poster presentations
- Taught the biology behind the overall lab goals, techniques used in the lab, and troubleshooting approaches

Guest Lecturer

(BIOL-322) Genetics and Molecular Biology
Indiana University Purdue University Indianapolis
Indianapolis, Indiana

Course Director: Dr. Kristin Chun
Lecture Title: Molecular Biology of Transcription and RNA Processing
Date: September 30, 2014

Teaching Assistant

(GRAD-716) Molecular Biology and Genetics
Indiana University School of Medicine
Indianapolis, Indiana
August 2011 – December 2011

- Tutored and led discussions with Ph.D. and M.D./Ph.D students for this semester-long class
- Duties included holding office hours for two hours every week, student-outreach for struggling students, and proctoring four examinations throughout the semester

SERVICE

Founder, Central Indiana Science Outreach

Indianapolis, Indiana
September 2014 – Present

- Recruited fellow researchers to form professional group focused on communicating the importance of science to the public
- Lead group members to build outreach networks, promote our activities, organize monthly planning meetings
- Curate public events hosted by local science organizations and promote on social media and website

Co-Founder & Co-Organizer, Student-Only Data Club, Indiana University School of Medicine

Indianapolis, Indiana
October 2013 – May 2015

- Initiated the first student-only journal club and research group in the Biochemistry and Molecular Biology Department and secured funding and support from department chair
- Promoted activities to reach an average participation rate of approximately 20 graduate students
- Encouraged professional development, public speaking skills, and camaraderie amongst students by scheduling two 30-minute presentations twice monthly
- Implemented social media to connect fellow students and encourage feedback and collaboration

Student Representative to the Faculty, Indiana University School of Medicine

Department of Biochemistry & Molecular Biology
Indianapolis, Indiana University
August 2012 – August 2013

- Organized student volunteers for departmental events such as the yearly picnic, holiday party, and research day poster presentations
- Communicated with students and faculty to address new ideas and implement new activities