

Université de Sherbrooke

**Régulation du transcriptome codant et non-codant chez *Schizosaccharomyces pombe*:
facteurs et mécanismes impliqués dans la maturation 3' des ARNs
et la terminaison de la transcription**

Par
Jean-François Lemay
Programme de Biochimie

Thèse présentée à la Faculté de médecine et des sciences de la santé
en vue de l'obtention du grade de *Philosophiae Doctor* (Ph.D.) en Biochimie

Sherbrooke, Québec, Canada
Décembre, 2016

Membres du jury d'évaluation
Professeur François Bachand, Ph.D. – Département de Biochimie
Professeur Simon Labbé, Ph.D. – Département de Biochimie
Professeur Sherif Abou Elela, Ph.D. – Département de Microbiologie et Infectiologie
Professeur Jacques Côté, Ph.D. – Centre de Recherche sur le Cancer,
Département de Biologie Moléculaire, Biochimie Médicale et Pathologie, Université Laval

© Jean-François Lemay, 2016

*Le succès n'est pas la clé du bonheur.
Le bonheur est la clé du succès.
Si vous aimez ce que vous faites, vous réussirez.
- Albert Schweitzer*

RÉSUMÉ

Régulation du transcriptome codant et non-codant chez *Schizosaccharomyces pombe*: facteurs et mécanismes impliqués dans la maturation 3' des ARNs et la terminaison de la transcription

Par
Jean-François Lemay
Programme de Biochimie

Thèse présentée à la Faculté de médecine et des sciences de la santé
en vue de l'obtention du diplôme de *Philosophiae Doctor* (Ph.D.) en Biochimie,
Faculté de médecine et des sciences de la santé, Université de Sherbrooke,
Sherbrooke, Québec, Canada, J1H 5N4

La synthèse d'un ARNm eucaryotique dépend d'une suite d'étapes qui inclut notamment l'ajout d'une queue poly(A) à son extrémité 3'. Au noyau, la queue poly(A) des ARNms est liée par PABPN1 (*poly(A)-binding protein nuclear 1*). PABPN1 fut notamment caractérisée, d'après des études *in vitro*, pour stimuler la réaction de polyadénylation en plus de contrôler la taille ultime des queues poly(A). Cela dit, la ou les fonction(s) biologique(s) de PABPN1 est/sont cependant largement méconnue(s). Chez *Schizosaccharomyces pombe* (*S. pombe*), Pab2 est l'orthologue présumé de PABPN1. Or, mes travaux indiquent que Pab2 est fonctionnellement différente de PABPN1 à l'égard de son rôle sur le processus général de polyadénylation. Ainsi, *in vivo*, l'absence de Pab2 entraîne l'expression et l'accumulation d'un groupe limité d'ARNs hyperadénylés parmi lesquels se trouvent de nombreux petits ARNs nucléolaires non-codants (*snoRNAs*) lesquels constituent normalement un groupe abondant d'ARN poly(A)-. Mes résultats supportent ainsi un mécanisme par lequel des *snoRNAs* immatures poly(A)+, sont convertis en une forme mature poly(A)- par le biais de Pab2 et de l'activité 3'→5' exoribonucléase de l'exosome à ARN. Ces observations sont inusitées dans la mesure où elles associent une fonction pour une PABP dans la maturation d'ARNs non-codants, contrairement à la notion que les PABPs travaillent exclusivement au niveau des ARNms, en plus de procurer une nouvelle perspective face au mécanisme de recrutement de l'exosome à ARN à des substrats poly(A)+.

La formation de l'extrémité 3' d'un ARN est un processus étroitement lié à la terminaison de sa transcription. Pour les gènes codants, la terminaison transcriptionnelle est initiée par le clivage endonucléolytique du pré-ARNm. Ce clivage génère une extrémité d'ARN 5' libre laquelle sera ciblée par une exoribonucléase 5'→3' afin de mener à bien l'éviction de l'ARNPII de la matrice d'ADN (terminaison transcriptionnelle de type *torpedo*). Au contraire, chez *Saccharomyces cerevisiae* (*S. cerevisiae*), la majorité des gènes non-codants, incluant les *snoRNAs*, dépendent plutôt du complexe NNS (Nrd1/Nab3/Sen1)

pour la terminaison de leur transcription. Cela dit, il est incertain si le complexe NNS est conservé chez d'autres espèces. À cet égard, mes travaux indiquent que *S. pombe* est dépourvu d'un mécanisme de terminaison de la transcription de type NNS. Seb1, l'orthologue présumé de Nrd1 chez *S. pombe*, s'associe plutôt à la machinerie de clivage et de polyadénylation et influence la sélection de site de polyadénylation à l'échelle du génome. Mes résultats supportent ainsi l'utilisation de la machinerie de maturation 3' des ARNms comme principal vecteur de terminaison transcriptionnelle chez *S. pombe* et identifient Seb1 comme un facteur clé de ce processus.

L'évènement transcriptionnel étant hautement complexe, des erreurs peuvent arriver de manière stochastique menant à l'accumulation d'ARNs aberrants potentiellement néfastes pour la cellule. Or, mes travaux ont mis en lumière un mécanisme de surveillance co-transcriptionnel des ARNs impliquant l'exosome à ARN et lié à la terminaison de la transcription. Pour ce faire, l'exosome à ARN promeut la terminaison transcriptionnelle via la dégradation d'une extrémité 3' libre d'ARN devenue émergente suite au recul de l'ARNPII le long de la matrice d'ADN (phénomène de *backtracking*). Mes résultats supportent ainsi une terminaison de la transcription de type *torpedo* inversé ($3' \rightarrow 5'$) réévaluant par la même occasion le concept voulant que la terminaison de la transcription s'effectue uniquement selon une orientation $5' \rightarrow 3'$.

Somme toute, mes travaux de doctorat auront permis d'identifier et de caractériser plus en détail les facteurs et mécanismes impliqués dans la maturation 3' et la terminaison de la transcription des gènes codants et non-codants chez l'organisme modèle *S. pombe*.

Mots-clés: *Schizosaccharomyces pombe*, PABPN1, Pab2, exosome à ARN, Seb1, maturation 3', polyadénylation alternative, terminaison transcriptionnelle.

TABLE DES MATIÈRES

Résumé	iii
Table des matières	v
Liste des figures	ix
Liste des tableaux	xi
Liste des abréviations	xii
Introduction	6
1. La levure <i>Schizosaccharomyces pombe</i> en tant qu'organisme modèle	6
2. L'expression génique	8
3. L'évènement transcriptionnel	9
3.1 L'ARN polymérase ADN-dépendante (ARNP)	9
3.1 Le domaine CTD de l'ARNPII	10
3.1 Les modifications post-traductionnelles du domaine CTD durant le cycle transcriptionnel et leur influence sur les évènements de maturation co-transcriptionnels de l'ARN	10
4. La maturation 3' de l'ARNm	14
4.1 Les éléments en <i>cis</i> nécessaires à la maturation 3' de l'ARNm	14
4.2 La reconnaissance des PAS par des facteurs protéiques chez les mammifères	16
4.2.1 Le complexe CPSF	16
4.2.2 Le complexe CstF	17
4.2.3 Le complexe CFI	18
4.2.3 Le complexe CFII	18
4.3 La reconnaissance des PAS par des facteurs protéiques chez la levure	19
5. Les protéines liant les queues poly(A) (PABP)	20
5.1 Les domaines protéiques de PABPN1	21
5.2 Fonction nucléaire canonique de PABPN1	22
5.3 Pab2: l'orthologue putatif de PABPN1 chez <i>S. pombe</i>	25
6. La terminaison de la transcription	25
6.1 La terminaison de la transcription des gènes codants	26

6.1.1	La terminaison de la transcription de type <i>torpedo</i>	26
6.1.2	La terminaison de la transcription de type allostérique	28
6.1.3	Unification des modèles de terminaison transcriptionnelle	29
6.1.4	L'association entre la pause transcriptionnelle et la terminaison de la transcription	32
6.2	La terminaison transcriptionnelle d'ARN non-codants chez la levure	33
6.2.1	La terminaison transcriptionnelle NNS-dépendante	33
7.	Les petits ARN nucléolaires (<i>snoRNAs</i>)	36
8.	La biogénèse des <i>snoRNAs</i>	39
9.	L'exosome à ARN	41
9.1	La structure de l'exosome à ARN	42
9.2	Les activités enzymatique de l'exosome à ARN	43
9.2.1	Dis3	44
9.2.2	Rrp6	45
9.3	Les isoformes de l'exosome à ARN chez la levure	45
9.3	Les voies de dégradation/maturation 3' → 5' de l'exosome à ARN nucléaire ..	45
10.	Les cofacteurs de l'exosome à ARN	48
11.	Les déterminants dictant le choix entre Dis3 et Rrp6	50
12.	Hypothèses et objectifs des travaux de recherche	50
	Résultats	53
	Article 1	53
	Contribution	54
	Résumé	55
	Abstract	56
	Introduction	57
	Results	60
	Discussion	78
	Experimental procedures	84
	Acknowledgments	86
	References	86
	Supplementary experimental procedures	92

Supplementary references	97
Supplementary tables	98
Article 2	102
Contribution	103
Résumé	104
Abstract	105
Introduction	106
Results	108
Discussion	133
Experimental procedures	137
Acknowledgments	139
References	139
Supplementary experimental procedures	145
Supplementary references	153
Supplementary tables	155
Article 3	173
Contribution	174
Résumé	175
Abstract	176
Introduction	177
Results	179
Discussion	201
Acknowledgments	205
Author contributions	205
References	205
Experimental procedures	209
Supplementary references	217
Supplementary tables	220
Discussion	222
Article 1	222
L'influence du type de polyadénylation sur l'expression génique	222

La polyadénylation lors de la biogénèse des <i>snoRNAs</i>	223
Pab2: recruteur ou activateur de l'exosome à ARN?	224
Une fonction gène-spécifique pour Pab2	226
Pab2 et ses partenaires dans la biogénèse des <i>snoRNAs</i>	227
Les mécanismes entourant la biogénèse des <i>snoRNAs</i> chez <i>S. pombe</i>	229
Article 2	231
Divergence fonctionnelle entre Nrd1 et Seb1	231
Simplification du mécanisme de maturation 3'/terminaison transcriptionnelle chez <i>S. pombe</i>	233
Les déterminants dictant le recrutement de Seb1 au complexe d'élongation	234
La relation entre Seb1 et la viabilité cellulaire	237
Le mécanisme d'action de Seb1 lors de l'évènement de maturation 3' de l'ARN	239
Une conservation de fonction chez les organismes supérieurs?	242
Article 3	242
L'exosome à ARN nucléaire : un complexe aux multiples fonctions	242
L'importance du <i>backtracking</i> de l'ARNPII lors de la terminaison de la transcription médiée par l'exosome à ARN	243
Les causes potentielles du <i>backtracking</i>	245
La conservation du mécanisme de terminaison de la transcription de type <i>torpedo</i> inversé	246
La présence d'un mécanisme de terminaison de la transcription de type «double <i>torpedo</i> »	247
L'intégration des travaux de thèse sur la conception actuelle de la terminaison transcriptionnelle chez les eucaryotes	249
Conclusion	251
Remerciements.....	253
Liste des références.....	254

LISTE DES FIGURES

Introduction

Figure 1. Schéma de la modulation des modifications post-traductionnelles du domaine CTD et des évènements de maturation co-transcriptionnels de l'ARN naissant au cours du cycle transcriptionnel chez la levure	11
Figure 2. Schéma des domaines protéiques présents chez PABPN1 et Pab2	22
Figure 3. Schéma du mécanisme de polyadénylation PABPN1-dépendant	24
Figure 4. Schéma de la terminaison transcriptionnelle de type <i>torpedo</i>	27
Figure 5. Schéma de la terminaison transcriptionnelle de type allostérique	30
Figure 6. Schéma de l'unification des mécanismes de terminaison transcriptionnelle <i>torpedo</i> et allostérique	31
Figure 7. Schéma de la terminaison transcriptionnelle NNS-dépendante	36
Figure 8. Schéma des différents types de <i>snoRNA</i>	39
Figure 9. Schéma décrivant la connexion entre la terminaison transcriptionnelle NNS-dépendante et la maturation/dégradation d'ARNnc	41
Figure 10. Schéma de l'architecture de l'exosome à ARN	43
Figure 11. Schéma des voies de dégradation/maturation 3'→5' de l'exosome à ARN nucléaire	47
Figure 12. Résumé général des travaux de thèse	252

Article 1

Figure 1. Gene-specific changes in expression level in the <i>pab2Δ</i> strain	61
Figure 2. Accumulation of 3'-extended polyadenylated snoRNAs in <i>pab2Δ</i> cells	62
Figure 3. 3'-extended snoRNAs accumulate in discrete foci in <i>pab2Δ</i> cells	68
Figure 4. Functional and physical associations between Pab2 and the nuclear exosome	69-70
Figure 5. Pab2 is required for Rrp6-dependent processing, but functions in a pathway distinct from the core exosome	73
Figure 6. Recruitment of Pab2 to the 3'-end of snoRNA genes	77
Figure 7. Model for the role of Pab2 in snoRNA synthesis	83
Figure S1-S3. Density of RNA Pol II along snoRNA genes in WT and <i>pab2Δ</i> cells	66
Figure S4. Overlap of upregulated genes from <i>pab2Δ</i> , <i>rrp6Δ</i> , and <i>dis3-54</i> cells	69
Figure S5. Rrp6 directly interacts with Pab2	71
Figure S6. Cid14-independent accumulation of nuclear poly(A) bodies in <i>pab2Δ</i> cells	74
Figure S7. Absence of 3'-extended snoRNA accumulation in <i>ski7Δ</i> cells	74
Figure S8. The H/ACA snoRNP dyskerin associates with polyadenylated H/ACA snoRNAs	81

Article 2

Figure 1. Transcription termination defects in <i>Seb1</i> -depleted cells	110
Figure 2. <i>Seb1</i> interacts with the 3' end processing machinery and is enriched at the 3' end of genes	115

Figure 3. Seb1 levels affect poly(A) site selection	118
Figure 4. Seb1 requires functional CID and RRM domains for accurate 3' end processing and transcription termination	122
Figure 5. Seb1 binds to GUA-containing motifs downstream of p(A) sites	126
Figure 6. Seb1 levels affect the co-transcriptional assembly of the cleavage/polyadenylation machinery	129
Figure 7. Transcription kinetics contributes to Seb1-dependent polyadenylation site selection	132
Figure S1. Read-through transcription by RNAPII in Seb1-deficient cells	111
Figure S2. Seb1 is not the functional ortholog of Nrd1p	112
Figure S3. Genome-wide distribution of Seb1 at the 3' end of RNAPII genes	116
Figure S4. Alternative p(A) sites (APA) utilization in Seb1-deficient cells	119
Figure S5. Transcription termination defects upon mutations of the CID and RRM domains of Seb1	124
Figure S6. Mapping Seb1-HTP-bound RNAs by the CRAC technic	127
Figure S7. 3' end processing factors are aberrantly recruited in Seb1-depleted condition	130

Article 3

Figure 1. Cells deficient in core exosome subunits, but not the TRAMP complex, accumulate read-through RNAs	182
Figure 2. Widespread production of 3'-extended transcripts in exosome-depleted cells ..	184
Figure 3. The core exosome promotes termination of RNAPII transcription	187
Figure 4. Transcription-dependent recruitment of core exosome subunits to genes	190
Figure 5. The 3'→5' exonucleolytic activity of Dis3 and the exosome central channel are required for transcription termination	193
Figure 6. Transcription termination by the exosome is mechanistically linked to RNAPII backtracking	198
Figure 7. Model of co-transcriptional RNA surveillance by the RNA exosome by competition with 3' end processing	204
Figure S1. Exosome depletion results in the production of read-through RNAs	180
Figure S2. Read-through RNAs detected in exosome-depleted cells are cleaved and polyadenylated downstream of motifs associated with 3' end processing and can result in the production of chimeric transcripts	181
Figure S3. The Rrp6 exoribonuclease does not extensively contribute to the suppression of read-through transcript	183
Figure S4. Read-through transcription by RNAPII in exosome-deficient cells	188
Figure S5. Expression and growth phenotypes of D166N (endo-), D516N (exo-), and D166N D516N (endo- exo-) versions of <i>S. pombe</i> Dis3	192
Figure S6. The central channel is essential for viability and snoRNA processing, but does not impair exosome assembly, in <i>S. pombe</i>	195
Figure S7. Transcription termination by the RNA exosome is mechanistically linked to RNAPII backtracking and pausing	200

LISTE DES TABLEAUX

Article 1

Supplementary Table 1. List of up- and down-regulated genes in <i>pab2Δ</i> cells	98
Supplementary Table 2. <i>S. pombe</i> strains used in this study	101

Article 2

Supplementary Table 1. List of proteins identified in the Seb1-HTP purification	155
Supplementary Table 2. List of Seb1-HTP -associated proteins with putative functions in 3' end processing and transcription termination of mRNAs	165
Supplementary Table 3. List of the top 100 genes displaying 3' UTR lengthening upon Seb1-depletion	166
Supplementary Table 4. List of the top 100 k-mer and their associated Z-score derived from pyMOTIF analysis of Seb1-HTP CRAC experiments	168
Supplementary Table 5. List of <i>S. pombe</i> strains used in this study	171
Supplementary Table 5. List of <i>S. cerevisiae</i> strains used in this study	172
Supplementary Table 5. List of plasmids used in this study	172

Article 3

Supplementary Table 1. <i>S. pombe</i> strains used in this study	220
Supplementary Table 2. Plasmids used in this study	221

LISTE DES ABRÉVIATIONS

3'READS	<i>3' region extraction and deep sequencing</i>
3' UTR(s)	<i>3' untranslated region(s)</i>
6-AU	6-azauracil
ADN	Acide désoxyribonucléique
APA	<i>Alternative polyadenylation</i>
ARN(s)	Acide(s) ribonucléique(s)
ARNm(s)	ARN(s) messenger(s)
ARNnc(s)	ARN non-codant(s)
ARNPI/II/III	ARN polymérase ADN-dépendante I/II/III
ARNr	ARN ribosomal
ARNsb	ARN simple brin
ARNdb	ARN double brins
ARNt	ARN de transfert
CFI/II	<i>Cleavage factor I/II</i>
CFIA/B	<i>Cleavage factor I A/B</i>
ChIP	<i>Chromatin immunoprecipitation</i>
CID	<i>CTD-interacting domain</i>
CPF	<i>Cleavage and polyadenylation factor</i>
CPSF	<i>Cleavage and polyadenylation specificity factor</i>
CRAC	<i>Cross-linking and cDNA analysis</i>
CstF	<i>Cleavage stimulation factor</i>
CTD	<i>Carboxy-terminal domain</i>
CUT(s)	<i>Cryptic unstable transcript(s)</i>
DSE(s)	<i>Downstream sequence element(s)</i>
<i>E. coli</i>	<i>Escherichia coli</i>
EE	<i>Efficiency element</i>
FISH	<i>Fluorescent in situ hybridization</i>

LCD	<i>Low complexity domain</i>
MTREC	<i>Mtl1-Red1 core</i>
NEXT	<i>Nuclear exosome targeting</i>
NLS	<i>Nuclear localization signal</i>
NNS	<i>Nrd1-Nab3-Sen1</i>
PAP	<i>Poly(A) polymerase</i>
Poly(A)	Succession de ribonucléotides de type adénosine (A)
PABP	<i>Poly(A)-binding protein</i>
PABPN/C	<i>Nuclear/cytoplasmic poly(A)-binding protein</i>
PAS	<i>Poly(A) site</i>
PE	<i>Positioning element</i>
PFI	<i>Polyadenylation factor I</i>
Pré-ARNm(s)	Précurseur(s) d'ARN(s) messenger(s)
qPCR	<i>Quantitative polymerase chain reaction</i>
RIP	<i>RNA immunoprecipitation</i>
RNB	<i>RNase II catalytic domain</i>
RRM	<i>RNA recognition motif</i>
SDE	<i>Site-determining element</i>
<i>snoRNA(s)</i>	<i>small nucleolar RNA(s)</i>
<i>snRNA(s)</i>	<i>small nuclear RNA(s)</i>
<i>S. pombe</i>	<i>Schizosaccharomyces pombe</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
TRAMP	<i>TRf-Air-Mtr4 Polyadenylation</i>
TSS	Site d'initiation de la transcription / <i>transcription start site</i>
WT	<i>Wild-type</i>

INTRODUCTION

1. La levure *Schizosaccharomyces pombe* en tant qu'organisme modèle

L'utilisation des levures comme vecteur de recherche pour l'identification et la caractérisation des processus fondamentaux de la cellule a été propulsée par les nombreux avantages qu'offrent ces micro-organismes. Au contraire des bactéries, tel qu'*Escherichia coli* (*E. coli*) qui fut l'un des premiers organismes modèles utilisés, les levures se distinguent par leur statut d'organisme unicellulaire eucaryotique. En plus d'accorder un avantage indéniable sur le plan technique en offrant la possibilité d'obtenir un nombre imposant d'individus dans un court laps de temps, au contraire d'organismes pluricellulaires, les levures, en tant qu'organisme eucaryotique, ont également l'avantage de permettre l'étude de processus conservés chez l'humain, mais absents chez les bactéries, tels la biogenèse d'organelles et l'organisation du cytosquelette, ainsi que l'étude de mécanismes divergents significativement de leur équivalent bactérien, tels que ceux liés à la transcription, la traduction ou encore la réplication d'ADN. Ces études ont particulièrement bénéficié de la capacité qu'ont les levures à subir des réarrangements génétiques lors de leur croissance végétative, notamment par le biais de la recombinaison homologue. L'habilité à recombiner de l'ADN exogène à des endroits spécifiques du génome a ainsi permis la génération d'un nombre considérable de souche mutante (délétion/insertion ou modification de gène ainsi que la création de fusion traductionnelle). De plus, la transition aisée entre les états d'haploïdie et de diploïdie, de même que l'aptitude à maintenir la présence de plasmide répliquatif autonome sont d'autres avantages qui ont favorisé l'essor des levures, principalement *S. cerevisiae* et *S. pombe*, en tant qu'organismes modèles.

S. pombe, une levure à fission d'après son mode de division cellulaire (scissiparité ou division binaire), fut initialement isolée de la bière de millet, un breuvage traditionnel communément retrouvé en Afrique. De nos jours, *S. pombe* est isolée d'une variété de sources naturelles et est présente mondialement (Jeffares *et al.*, 2015). L'étendue et l'importance limitée des applications associées à *S. pombe*, comparativement à *S. cerevisiae*, ont toutefois

conféré à cette dernière une certaine préférence au sein de la communauté scientifique. L'utilisation de *S. pombe* en recherche a néanmoins permis d'importantes avancées notamment dans les domaines du cycle cellulaire et de la dynamique des chromosomes ainsi que sur l'influence de l'épigénétique sur le contrôle de l'expression génique (Hoffman *et al.*, 2015). Son importance en recherche est aussi soulignée par un partage de caractéristiques biologiques plus important avec les cellules mammifères qu'avec *S. cerevisiae*. *S. pombe* possède notamment de larges centromères modulaires composés majoritairement d'ADN répété qui s'étendent sur plusieurs milliers de paires de bases, rappelant ceux d'organismes supérieurs. En comparaison, les centromères chez *S. cerevisiae* ne s'étendent que sur quelques paires de bases et sont dépourvus d'ADN répété (Wood *et al.*, 2002). *S. pombe* a de plus des origines de réplication et un contenu intronique s'approchant davantage de ceux retrouvés chez les eucaryotes supérieurs que de ceux de chez *S. cerevisiae* (Mojardin *et al.*, 2013, Wood *et al.*, 2002). *S. pombe* possède également quelques 338 protéines conservées chez les vertébrés, mais absentes chez *S. cerevisiae*, parmi lesquelles se trouvent notamment les composantes de la machinerie d'interférence à l'ARN (Aravind *et al.*, 2000, Wood *et al.*, 2012)¹. Ces gènes font partie des 66% (3399/5138) possédant une équivalence chez les vertébrés (Wood *et al.*, 2012)². De manière intéressante, parmi ces gènes, 501 sont présumés associés à des maladies humaines (Wood *et al.*, 2012)³ et de ce nombre 50 gènes sont uniquement présents chez *S. pombe*, n'étant pas conservés chez *S. cerevisiae* (Wood *et al.*, 2012)⁴. Avec en prime la complétion du séquençage et de l'annotation du génome (Dutrow *et al.*, 2008, Rhind *et al.*, 2011, Wilhelm *et al.*, 2008) ainsi qu'un accès grandissant à des études transcriptomiques, protéomiques, génétiques et phénotypiques à large échelle (caractérisation de la localisation cellulaire de l'*ORFeome* (Matsuyama *et al.*, 2006), détermination des gènes essentiels pour la viabilité cellulaire (Kim *et al.*, 2010), quantification absolue du protéome, du phosphoprotéome et du transcriptome (Carpy *et al.*,

¹ <http://www.pombase.org/spombe/query/builder>. Éléments de recherche: PBO:0011070 (*conserved in vertebrates*) et PBO:0000055 (*no apparent S. cerevisiae ortholog*). En date d'avril 2016.

² <http://www.pombase.org/spombe/query/builder>. Éléments de recherche: *Feature Type (protein coding)* et PBO:0011070 (*conserved in vertebrates*). En date d'avril 2016.

³ <http://www.pombase.org/spombe/query/builder>. Éléments de recherche: *Feature Type (protein coding)*, PBO:5000000 (*disease associated*) et PBO:0011070 (*conserved in vertebrates*). En date d'avril 2016.

⁴ <http://www.pombase.org/spombe/query/builder>. Éléments de recherche: *Feature Type (protein coding)*, PBO:5000000 (*disease associated*), PBO:0011070 (*conserved in vertebrates*) et PBO:0000055 (*no apparent S. cerevisiae ortholog*). En date d'avril 2016.

2014, Marguerat *et al.*, 2012), identification des sites de clivage et de polyadénylation (Mata, 2013, Schlackow *et al.*, 2013), détermination du *translatome* lors de la différenciation sexuelle (Duncan et Mata, 2014), caractérisation du réseau d'interaction protéique (Vo *et al.*, 2016) et caractérisation des paramètres cinétiques du métabolisme d'ARN (Eser *et al.*, 2016)) *S. pombe* peut donc être considéré comme un système de prédilection pour l'étude des principes de bases de la cellule et un vecteur de connaissances important pour une meilleure compréhension d'organismes plus complexes tels que les humains.

2. L'expression génique

L'ensemble des instructions nécessaires au maintien des activités cellulaires se trouve inscrit à l'intérieur du code génétique. L'expression génique renvoie donc au processus global par lequel l'information contenu à l'intérieur d'une région définie d'ADN, appelée gène, est utilisée pour la synthèse de matériel biochimique. Ce produit peut être soit une protéine ou encore un ARN selon le type de gène étudié (codant versus non-codant respectivement). Les processus qui dictent l'expression génique sont particulièrement complexes et, à bien des égards, l'activité génique, définie par la quantité du produit d'un gène donné, est le reflet des fluctuations temporelles des taux relatifs d'évènements de synthèse et de dégradation. Virtuellement n'importe quelle étape du processus d'expression génique peut être modulée. Cette régulation peut avoir lieu lors du processus transcriptionnel ou traductionnel, mais également lors des évènements dictant la maturation co- et post-transcriptionnelle de l'ARN, le transit intracellulaire de l'ARN ou encore l'ajout de modifications post-traductionnelles aux protéines (Komili et Silver, 2008). Ces points de régulation sont d'autant plus importants qu'ils servent, en outre, à assurer l'intégrité du contenu ribonucléique de la cellule, mais aussi à conférer une capacité d'adaptation face aux changements environnementaux.

3. L'évènement transcriptionnel

Les gènes eucaryotiques sont transcrits par l'entremise d'un cycle composé de trois phases séquentielles, mais interconnectées, débutant avec la phase d'initiation et se poursuivant respectivement avec les phases d'élongation et de terminaison de la transcription. Chacune de ces phases dépend d'un nombre imposant de facteurs, soulignant ainsi la nature complexe et l'important potentiel de régulation qu'offre l'évènement transcriptionnel dans son ensemble. Cela dit, la protéine centrale du processus transcriptionnel est l'ARN polymérase ADN-dépendante.

3.1. L'ARN polymérase ADN-dépendante (ARNP)

La polymérisation de ribonucléotides libres en une chaîne d'acide ribonucléique chez les eucaryotes est sous le contrôle de trois ARNs polymérases ADN-dépendantes nucléaires, l'ARNPI, II et III, chacune ayant un rôle primordial au bon fonctionnement de la cellule. L'ARNPI transcrit la plupart des ARNs ribosomiaux (ARNr) (5.8S, 18S et 25S) (Russell et Zomerdijk, 2005) tandis que l'ARNPIII transcrits les ARNs de transfert (ARNt), l'ARNr 5S, le petit ARN nucléaire U6 ainsi qu'une variété d'autres ARNs de petites tailles (Dieci *et al.*, 2013). Combiné, la transcription médiée par l'ARNPI et III équivaut à plus de la moitié du débit transcriptionnel de la cellule (Warner, 1999). Malgré une contribution absolue moindre, l'activité transcriptionnelle de l'ARNPII reste la plus variée, allant de la transcription des gènes codant pour des protéines à la transcription d'un nombre important d'ARNs non-codants (ARNncs) tels que les petits ARNs nucléaires et nucléolaires (*snRNAs* et *snoRNAs* respectivement) de même que les transcrits d'origine hétérochromatique ou cryptique (intergénique, antisense).

Les ARNPs sont des enzymes aux sous-unités multiples. Celles-ci sont structurellement liées par la présence d'un noyau multiprotéique conservé auquel vient s'associer des sous-unités additionnelles (Cramer *et al.*, 2008). L'ARNPII est composée de 12 polypeptides (Rpb1 à Rpb12) dont 10 s'assemblent en un noyau auquel vient se greffer en périphérie un sous-complexe formé de l'hétérodimère Rpb4/7 (Armache *et al.*, 2005). Des 12 sous-unités

formant l'ARNPII, 5 sont communes aux 3 ARNPs alors que les 7 autres possèdent des similarités de séquence et de structure à des sous-unités présentes chez l'ARNPI et III. L'ARNPII se distingue de l'ARNPI et III par la présence, au niveau de sa sous-unité Rpb1, d'un domaine carboxy-terminal (CTD) flexible situé à proximité de l'emplacement où l'ARN naissant émerge de l'ARNPII (Allison *et al.*, 1985, Chen *et al.*, 2009, Corden *et al.*, 1985, Cramer *et al.*, 2001, Spahr *et al.*, 2009).

3.2. Le domaine CTD de l'ARNPII

Le domaine CTD de l'ARNPII est présent chez un large éventail d'organismes (Liu *et al.*, 2010). La grande majorité de ce domaine est composé de répétitions de l'heptapeptide $Y^1S^2P^3T^4S^5P^6S^7$ dont le nombre varie parmi les taxa allant de 26 chez *S. cerevisiae* (Allison *et al.*, 1985) à 29 chez *S. pombe* (Azuma *et al.*, 1991) et jusqu'à 52 chez les mammifères (Corden *et al.*, 1985).

Le domaine CTD est essentiel à la viabilité cellulaire. Toutefois, chez les levures et les cellules mammifères un nombre limité d'heptades peut être enlevé sans conséquences phénotypiques notables. Ainsi, chez *S. cerevisiae*, 10 des 26 répétitions sont nécessaires pour l'obtention d'une croissance similaire à une souche sauvage (WT) (West et Corden, 1995) alors que ce nombre est respectivement de 16 et 29 chez *S. pombe* et la souris (Bartolomei *et al.*, 1988, Schneider *et al.*, 2010). Cette capacité à tronquer le domaine CTD illustre la redondance fonctionnelle des répétitions de ce domaine.

3.3. Les modifications post-traductionnelles du domaine CTD durant le cycle transcriptionnel et leur influence sur les évènements de maturation co-transcriptionnels de l'ARN

Durant l'évènement transcriptionnel, le domaine CTD de l'ARNPII sert au recrutement d'une gamme diversifiée de facteurs (Hirose et Manley, 1998, Hsin et Manley, 2012, Jeronimo *et al.*, 2013, McCracken *et al.*, 1997). Cette capacité à interagir avec des facteurs donnés nécessite des modifications post-traductionnelles spécifiques du domaine CTD

lesquelles sont dynamiquement modulées en fonction de la progression du cycle transcriptionnel, un concept connu comme le «code» du domaine CTD (Buratowski, 2003). Des 7 acides aminés composant un heptade, 5 peuvent être phosphorylés (Y^1 , S^2 , T^4 , S^5 et S^7) tandis que les prolines (P^3 et P^6) peuvent isomériser entre une conformation *cis* et *trans* (Heidemann *et al.*, 2013). De nombreuses études ont répertorié les transitions des marques de phosphorylation durant l'évènement transcriptionnel chez l'humain et *S. cerevisiae* (Bataille *et al.*, 2012, Kim *et al.*, 2010, Komarnitsky *et al.*, 2000, Mayer *et al.*, 2012, Mayer *et al.*, 2010, Tietjen *et al.*, 2010) (Figure 1).

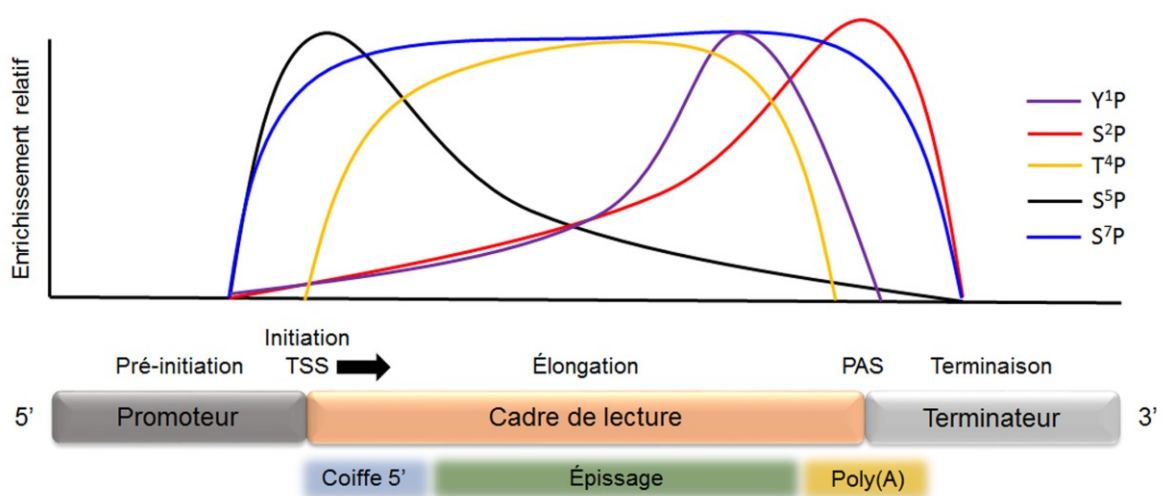


Figure 1. Schéma de la modulation des modifications post-traductionnelles du domaine CTD et des évènements de maturation co-transcriptionnels de l'ARN naissant au cours du cycle transcriptionnel chez *S. cerevisiae*.

Au promoteur des gènes, l'ARNPII est sous forme hypo-phosphorylée (pré-initiation), mais est rapidement phosphorylée au niveau de la S^5 et S^7 au cours de la phase d'initiation de la transcription. La phosphorylation de la S^5 est primordiale pour l'ajout de la structure coiffe à l'extrémité 5' du pré-ARNm. Durant la phase d'élongation, la phosphorylation de la S^5 diminue graduellement au profit de la phosphorylation de la S^2 et de la Y^1 . Cette transition est associée au recrutement de facteurs stimulant l'élongation de la transcription ainsi que des facteurs nécessaires à l'épissage du pré-ARNm. Au contraire de la phosphorylation de la Y^1 qui chute drastiquement en amont du site de polyadénylation (PAS), la phosphorylation de la S^2 culmine au PAS ce qui coïncide avec le recrutement de facteurs impliqués au niveau du clivage et de la polyadénylation du pré-ARNm et la terminaison de la transcription. Le profil de la phosphorylation de la T^4 est pour sa part largement confiné au cadre de lecture alors que celui de la S^7 est prévalent sur l'ensemble de l'unité transcriptionnelle. Durant la phase de terminaison de la transcription, le domaine CTD revient à un état hypo-phosphorylé

afin de débiter un nouveau cycle transcriptionnel. Les patrons de distribution des modifications du domaine CTD sont dérivés d'expériences de ChIP lesquelles utilisent des anticorps modification-spécifique. Ces patrons représentent l'enrichissement relatif pour chacune des modifications et ne peuvent donc pas servir de base de comparaison entre celles-ci. À cet égard, le niveau de phosphorylation de la Y¹, de la T⁴ et de la S⁷ est largement sous-représenté par rapport à celui de la S² et S⁵ lorsque quantifié par spectrométrie de masse. TSS : site d'initiation de la transcription. Le schéma n'est pas à l'échelle.

Durant la phase d'initiation de la transcription, une version hypo-phosphorylée de l'ARNPII est recrutée au promoteur des gènes et s'assemble avec des facteurs généraux de transcription afin de former un complexe de pré-initiation (Grunberg et Hahn, 2013). Les premières marques de phosphorylation à apparaître se produisent au niveau de la S⁵ et de la S⁷. Celles-ci résultent de l'action de la kinase cycline-dépendante Cdk7 (Kin28 et Mcs6 chez *S. cerevisiae* et *S. pombe* respectivement), une sous-unité du facteur générale de transcription TFIIF (Akhtar *et al.*, 2009). La principale fonction de la phosphorylation de la S⁵ consiste à recruter la machinerie enzymatique nécessaire à l'ajout de la structure coiffe à l'extrémité 5' du pré-ARNm naissant (Cho *et al.*, 1997, Ho et Shuman, 1999, McCracken *et al.*, 1997, Yue *et al.*, 1997). À cet égard, chez *S. pombe*, la substitution de l'ensemble des S⁵ en alanine (S⁵→A⁵) entraîne un phénotype d'invivabilité. Or, ce phénotype peut être contourné par le truchement d'une liaison covalente entre le domaine CTD et le complexe enzymatique nécessaire à l'ajout de la structure coiffe (Schwer et Shuman, 2011). La phosphorylation de la S⁵ participe également au recrutement des facteurs nécessaires pour la terminaison des gènes non-codants chez *S. cerevisiae* (Kubicek *et al.*, 2012, Vasiljeva *et al.*, 2008) (voir la section 6.2.1). La présence de cette modification est intimement liée à la distance parcourue par l'ARNPII en cours d'élongation comme l'est d'ailleurs l'ensemble des modifications du domaine CTD. La présence de S⁷ phosphorylé est, quant à elle, relativement stable sur l'ensemble de l'unité transcriptionnelle (Bataille *et al.*, 2012, Kim *et al.*, 2010, Mayer *et al.*, 2010, Tietjen *et al.*, 2010). Au contraire de la S⁵, l'implication de la phosphorylation de la S⁷ sur les évènements de maturation co-transcriptionnels de l'ARN chez *S. pombe* et *S. cerevisiae* sont inexistantes. Par contre, chez l'humain, cette modification est liée à la maturation des *snRNAs* par le complexe *Integrator* (Egloff *et al.*, 2007, Egloff *et al.*, 2010). Pour sa part, la phosphorylation de la Y¹ s'accroît graduellement le long de l'unité transcriptionnelle pour s'estomper juste avant l'atteinte du PAS. Ce profil est important afin

de restreindre le recrutement des facteurs liés aux événements de maturation 3' et de terminaison de la transcription à la région immédiate du PAS (Mayer *et al.*, 2012). Le domaine CTD est également phosphorylé au niveau de la T⁴. Chez *S. cerevisiae*, cette modification est limitée aux cadres de lecture alors qu'elle est principalement localisée juste en amont du PAS chez l'humain (Hintermair *et al.*, 2012, Mayer *et al.*, 2012). La phosphorylation de la T⁴ est notamment nécessaire à la maturation 3' des ARNms d'histones chez des cellules de poulet (Hsin *et al.*, 2011), à l'élongation de la transcription chez l'humain (Hintermair *et al.*, 2012) ainsi qu'au remodelage de la chromatine chez certains gènes chez *S. cerevisiae* (Rosonina *et al.*, 2014). Suite à la phosphorylation de la S⁵ apparaît celle de la S². La phosphorylation de la S² s'accroît graduellement à partir du site d'initiation de la transcription (TSS) pour culminer dans l'environnement du PAS (Bataille *et al.*, 2012, Kim *et al.*, 2010, Mayer *et al.*, 2010, Tietjen *et al.*, 2010). L'émergence de cette modification concorde avec l'entrée de l'ARNPII en un mode d'élongation (Bowman et Kelly, 2014). Chez les levures, la présence de S² phosphorylée dépend de l'action de deux kinases: une majeure, Ctk1/Lsk1 (*S. cerevisiae/S. pombe*), et une mineure, Bur1/Cdk9 (Qiu *et al.*, 2009). Celles-ci coordonnent le recrutement des facteurs favorisant l'élongation de la transcription, l'épissage, l'export des ARNms au cytoplasme ainsi que la maturation 3' du pré-ARNm et la terminaison de la transcription (Jeronimo *et al.*, 2013).

Le «code» du domaine CTD apparaît cependant passablement moins complexe qu'envisagé. Ainsi, chez la levure et l'humain, des analyses protéomiques effectuées sur une version modifiée du CTD permettant son analyse par spectrométrie de masse ont révélé que le domaine CTD est principalement phosphorylé en S² et S⁵ avec peu ou pas de phosphorylation au niveau de la Y¹, T⁴ et S⁷ (Schuller *et al.*, 2016, Suh *et al.*, 2016). L'utilisation de cette approche a notamment l'avantage de ne pas dépendre d'anticorps modification-spécifique dont l'affinité pourrait être influencée positivement ou négativement par des modifications environnantes. Ces données n'excluent toutefois pas un rôle pour la phosphorylation de la Y¹, T⁴ et S⁷ sur le processus transcriptionnel et/ou les événements de maturation co-transcriptionnels de l'ARN, mais suggèrent que la modification post-traductionnelle de ces acides aminés est plus dynamique/transitoire ou encore limitée à des classes particulières de gènes.

4. La maturation 3' de l'ARNm

La formation de l'extrémité 3' des précurseurs d'ARNms est une étape essentielle dans le processus d'expression génique. La maturation 3' des pré-ARNms eucaryotiques, à l'exception des ARNms d'histones, se caractérise par une réaction en deux étapes comprenant le clivage du pré-ARNm naissant par une endonucléase et la synthèse d'une queue poly(A) par une poly(A) polymérase (PAP) au niveau de l'extrémité 3' du produit de clivage en amont. En apparence simple, l'évènement de maturation 3' implique une combinaison d'éléments en *cis* au niveau du transcrit naissant ainsi qu'une machinerie protéique formée de plusieurs dizaines de sous-unités.

4.1. Les éléments en *cis* nécessaires à la maturation 3' de l'ARNm

Pour que la maturation 3' s'exécute correctement, le pré-ARNm doit être pourvu d'éléments en *cis* afin de guider le positionnement adéquatement de protéines effectrices. L'ensemble de ces éléments, présents à la fois en amont et en aval du site de clivage, définit le PAS (Tian et Graber, 2012). Chez les mammifères, les PAS sont composés de plusieurs éléments en *cis* distincts. De ces éléments, 3 se trouvent en amont du site de coupure. Le premier élément est formé d'un motif hexamérique, principalement AATAAA ou ATTTAA, positionné environ -10 à -35 nucléotides du site de coupure (Beaudoing *et al.*, 2000). Le second élément est situé entre les nucléotides -40 à -100 et se compose de plusieurs répétitions du motif TGTA. Finalement, le troisième élément est caractérisé par une composition nucléotidique riche en T et est positionné entre l'élément hexamérique et le site de clivage. Des éléments en *cis* sont également trouvés en aval du site de coupure. Parmi ceux-ci, se trouvent les éléments DSEs (pour *downstream elements*) composés de séquences riches en T et en GT. Ceux-ci sont préférentiellement enrichis dans une fenêtre d'environ 40 nucléotides suivant le site de coupure avec une tendance pour les séquences riches en GT à y être positionnés plus près. Les PAS des mammifères possèdent également un élément en *cis* riche en G à la suite des éléments DSEs (Hu *et al.*, 2005).

La grande majorité des connaissances entourant la composition des PAS chez les levures provient d'études effectuées chez *S. cerevisiae*. Le modèle actuel inclue 4 éléments en *cis*: un élément EE, PE ainsi que deux éléments riches en T. L'élément EE (pour *efficiency element*) se situe en amont du site de clivage et est caractérisé par un contexte nucléotidique riche en TA (TATATA étant la séquence optimale) (Russo *et al.*, 1991). Ce motif sert à accroître l'efficacité des éléments situés en aval. Parmi ceux-ci se trouvent, l'élément PE (pour *positioning element*) (Russo *et al.*, 1993). Le motif PE est équivalent au motif hexamérique des PAS chez les mammifères et sert donc à définir l'emplacement du site de clivage. AATAAA et AAAAAA sont les motifs optimaux de l'élément PE (Guo et Sherman, 1995). Finalement, deux éléments riches en T ont été identifiés de part et d'autre du site de clivage (Dichtl et Keller, 2001, Graber *et al.*, 1999). Cette organisation d'éléments en *cis* chez la levure (EE-PE-T-p(A)-T) rappelle celle présente chez les mammifères (T-hexamère-T-p(A)-T-G). Les éléments en *cis* chez la levure sont toutefois plus sujets à être dégénérés (Graber *et al.*, 1999, Guo et Sherman, 1995).

Chez *S. pombe*, les éléments en *cis* ont été principalement caractérisés au niveau du gène *ura4*. Le PAS du gène *ura4* est complexe et s'étend sur plusieurs centaines de nucléotides. Il comporte un élément SDE (pour *site-determining element*), positionnée en amont du site de clivage. L'élément SDE possède notamment une séquence agissant comme le motif PE de chez *S. cerevisiae* ou le motif hexamérique de mammifères. Un deuxième élément, l'élément EE (pour *efficiency element*), agit de manière à augmenter considérablement l'efficacité du processus de clivage (Humphrey *et al.*, 1994). On trouve également un élément DSE (pour *downstream sequence element*) en aval du site de coupure. L'ensemble de ces éléments sont notamment nécessaires afin d'induire un état de pause au niveau du complexe d'élongation de l'ARNPII de manière à favoriser la maturation 3' de l'ARN et la terminaison de la transcription (Aranda et Proudfoot, 1999, Birse *et al.*, 1997). En général, les éléments en *cis* chez *S. pombe* exhibent une organisation semblable à celle observée chez *S. cerevisiae* et les mammifères. Ces éléments sont également fonctionnellement similaires dans la mesure où un PAS de chez *S. pombe* est reconnu par la machinerie de clivage et de polyadénylation présente chez *S. cerevisiae* et vice versa (Humphrey *et al.*, 1991). Une telle interchangeabilité n'est toutefois pas observable avec les PAS des mammifères et suggère une reconnaissance

différentielle des motifs en *cis* par les facteurs protéiques impliqués dans la maturation 3' entre les levures et les mammifères.

4.2. La reconnaissance des PAS par des facteurs protéiques chez les mammifères

La maturation 3' d'un ARNm est accomplie par des actions coopératives entre plusieurs complexes protéiques. Chez les mammifères, 4 complexes multiprotéiques forment l'unité de base de la machinerie de clivage et de polyadénylation soit les complexes CPSF (pour *cleavage and polyadenylation specificity factor*), CstF (pour *cleavage stimulation factor*), CFI et CFII (pour *cleavage factor* I et II respectivement) (Takagaki *et al.*, 1989). Tous ces complexes sont indispensables pour une réaction de clivage efficace. Ces 4 complexes interagissent également avec plusieurs autres protéines. Des quelques 85 protéines identifiées lors d'une étude protéomique effectuée chez des cellules humaines, plusieurs sont des composantes intégrales de la machinerie de maturation 3', telles que l'ARNPII, PABPN1 ou encore la PAP, alors que d'autres ont plutôt des fonctions de régulation ou servent d'intermédiaires entre la maturation 3' et d'autres processus cellulaires (Shi *et al.*, 2009).

4.2.1. Le complexe CPSF

CPSF est un complexe multiprotéique recruté au niveau du domaine CTD de l'ARNPII et est composé des sous-unités CPSF-160, CPSF-100, CPSF-73, CPSF-30, Wdr33 et Fip1 (Bienroth *et al.*, 1991, Kaufmann *et al.*, 2004, McCracken *et al.*, 1997, Murthy et Manley, 1992, Shi *et al.*, 2009). Deux de ces protéines, Wdr33 et CPSF-30, lient directement, d'après des études *in vitro* et *in vivo*, une étroite région d'ARN couvrant le motif hexamérique du PAS (Chan *et al.*, 2014, Schonemann *et al.*, 2014). Ces interactions permettent ainsi d'ancrer le complexe CPSF à l'ARN naissant et de fournir une spécificité à la réaction de maturation 3'. Outre son affinité pour l'ARN, CPSF-30 interagit également avec l'ARNPII influençant l'association du complexe CPSF à l'ARNPII durant l'événement transcriptionnel (Nag *et al.*, 2007). Tout comme CPSF-30, la plus imposante sous-unité du complexe CPSF, CPSF-160, est également impliquée dans des interactions type protéine-ARN et protéine-protéine. CPSF-160 interagit notamment avec l'ARN sur une région diffuse en amont du site de

clivage (Gilmartin *et al.*, 1995, Martin *et al.*, 2012, Murthy et Manley, 1995) de manière à stabiliser le complexe CPSF au niveau du pré-ARNm. CPSF-160 établie de plus des interactions avec la composante CstF-77 du complexe CstF ainsi qu'avec la PAP (Murthy et Manley, 1995). Fip1 contribue également à la liaison du complexe CPSF au PAS. L'interaction de Fip1 avec le pré-ARNm s'effectue notamment au niveau de séquences riches en U situées en amont du motif hexamérique (Kaufmann *et al.*, 2004, Lackford *et al.*, 2014, Martin *et al.*, 2012). Fip1 établie aussi des interactions avec d'autres composantes de la machinerie de clivage et de polyadénylation, interagissant avec CPSF-30, CPSF-160, CstF-77 ainsi qu'avec la PAP (Kaufmann *et al.*, 2004, Murthy et Manley, 1995). Au-delà de la liaison au PAS, le complexe CPSF catalyse le clivage du pré-ARNm au site d'addition de la queue poly(A) par le biais de sa sous-unité CPSF-73 (Mandel *et al.*, 2006, Ryan *et al.*, 2004), une protéine appartenant à la famille d'hydrolase métal-dépendante à domaine β -CASP (Callebaut *et al.*, 2002). CPSF-100, quant à elle, possède une séquence protéique hautement similaire à celle de CPSF-73 (Callebaut *et al.*, 2002). Tout comme CPSF-73, CPSF-100 possède un domaine β -CASP. Toutefois, les résidus du domaine β -CASP impliqués dans la liaison aux ions métalliques sont absents chez CPSF-100 privant celle-ci de fonction enzymatique (Callebaut *et al.*, 2002, Mandel *et al.*, 2006). CPSF-100 possède également un domaine à son extrémité carboxy permettant son hétérodimérisation avec CPSF-73 (Dominski *et al.*, 2005). Le rôle de CPSF-100 dans le contexte de la machinerie de clivage et de polyadénylation semble ainsi plus lié à une fonction de régulation de l'activité enzymatique de CPSF-73 via son interaction avec cette dernière.

4.2.2. Le complexe CstF

CstF est un complexe formé de trois sous-unités: CstF-77, CstF-64 et CstF-50 (Gilmartin et Nevins, 1991, Takagaki *et al.*, 1990). Son rôle au niveau de la maturation 3' est le reflet à la fois de son interaction avec le domaine CTD de l'ARNPII (Fong et Bentley, 2001, McCracken *et al.*, 1997), le complexe CPSF (Gilmartin et Nevins, 1991, Wilusz *et al.*, 1990) ainsi qu'avec le pré-ARNm via sa liaison aux éléments DSEs. L'interaction du complexe CstF avec le pré-ARNm s'effectue via le domaine de liaison à l'ARN (RRM : *RNA recognition motif*) de CstF-64. Celui-ci lie préférentiellement des séquences riches en U, mais

peu également lier des séquences riches en GU (MacDonald *et al.*, 1994, Martin *et al.*, 2012, Takagaki et Manley, 1997). Pour sa part, CstF-77 permet d'unir CstF-50 et CstF-64 au sein d'un même complexe (Takagaki et Manley, 2000). Quant à la sous-unité CstF-50, celle-ci est uniquement présente chez les mammifères et ses fonctions à l'intérieur du complexe CstF sont peu connues.

4.2.3. Le complexe CFI

Le complexe CFI existe sous la forme d'un hétérotétramère formé de l'agencement de deux dimères à la composition protéique variable. Chaque dimère se compose d'une sous-unité invariable, CFI-25, à laquelle une de deux protéines hautement similaires CFI-68 ou CFI-59 vient se lier (Ruegsegger *et al.*, 1996). Ce complexe est présent uniquement chez les métazoaires. Des expériences de sélection *in vitro* de substrat d'ARN (SELEX) ont permis d'identifier comme motif de liaison pour le complexe CFI la séquence UGUAN (Brown et Gilmartin, 2003), une séquence qui a depuis été validée par des approches structurales et de liaisons *in vivo* (Martin *et al.*, 2012, Yang *et al.*, 2010). Au niveau de la reconnaissance globale du PAS, CFI joue un rôle important en favorisant le recrutement et la stabilisation du complexe CPSF sur le pré-ARNm (Ruegsegger *et al.*, 1996, Venkataraman *et al.*, 2005).

4.2.4. Le complexe CFII

Parmi tous les facteurs formant la machinerie de base de clivage et polyadénylation chez les mammifères, le complexe CFII est sans contredit le complexe multiprotéique le moins bien caractérisé. Deux protéines composent le complexe CFII chez les mammifères : Pcf11 et Clp1 (de Vries *et al.*, 2000). CFII interagit toutefois avec plusieurs autres protéines faisant de ce complexe un point central pour la coordination de nombreux événements cellulaires (de Vries *et al.*, 2000). Pcf11 a un rôle au niveau de la promotion de la terminaison de la transcription tandis que Clp1 a un rôle d'échafaudage en permettant l'interaction aux complexes CPSF et CFI (de Vries *et al.*, 2000, West et Proudfoot, 2008).

4.3. La reconnaissance des PAS par des facteurs protéiques chez la levure

La grande majorité des composantes protéiques formant la machinerie de clivage et de polyadénylation chez les mammifères est présente chez la levure. L'arrangement de ces facteurs en structures multiprotéiques est toutefois légèrement différent. Chez *S. cerevisiae*, la machinerie de maturation 3' est formée des complexes CPF (pour *cleavage and polyadenylation factor*), CFIA et CFIB (pour *cleavage factor A/B*). Le complexe CPF est lui-même formé de deux sous-complexes, CFII (pour *cleavage factor II*) et PFI (pour *polyadenylation factor I*). Le complexe CFII contient notamment les sous-unités orthologues aux facteurs trouvés au sein du complexe CPSF des mammifères (CPSF-160/Cft1, CPSF-100/Cft2, CPSF-73/Ysh1) à l'exception de CPSF-30 (Yth1), Fip1 (Fip1) et Wdr33 (Psf2), lesquels font partis du complexe PFI. Pour sa part, CFIA est composé de facteurs orthologues à ceux des complexes CstF (CstF77/Rna15, CstF-64/Rna14) et CFII des mammifères (Pcf11/Pcf11 et Clp1/Clp1). CFIB, formé uniquement de Hrp1, est, quant à lui, unique aux levures (Mandel *et al.*, 2008, Xiang *et al.*, 2014). Plusieurs de ces protéines sont connus pour lier l'ARN et le domaine CTD de l'ARNPII (Barilla *et al.*, 2001, Dichtl *et al.*, 2002, Kyburz *et al.*, 2003) et ont une fonction équivalente à leur contrepartie mammifère. Ysh1 est notamment responsable du clivage endonucléolytique du pré-ARNm (Ryan *et al.*, 2004). Cft1 et Cft2 lient toutes deux l'ARN à proximité du site de coupure via des séquences riches en U (Dichtl *et al.*, 2002, Dichtl et Keller, 2001). Une telle préférence de liaison est également observée pour Yth1 (Barabino *et al.*, 2000). Toutefois, une étude d'identification de site de liaison à l'ARN *in vivo* indique un patron de liaison distinctif pour Yth1 par rapport à d'autres sous-unités du complexe CPF (Baejen *et al.*, 2014). En se liant environ 17 nucléotides avant le site de coupure, Yth1 lierait l'environnement immédiat de l'élément PE ce qui serait en accord avec sa fonction connue chez les mammifères. Fip1 des levures, mammifères ou d'autres espèces partagent une similarité en ce qui attrait à l'organisation de leurs domaines protéiques. Toutefois, chez *S. cerevisiae*, Fip1 ne possède pas le domaine RD ni le domaine riche en arginine présent à l'extrémité carboxy de la version mammifère lesquels sont nécessaires à la liaison à l'ARN (Kaufmann *et al.*, 2004). Fip1 semble donc avoir évolué de manière à perdre sa capacité à lier l'ARN au profit d'interaction protéine-protéine. Or, Fip1 fut notamment identifiée suite à son interaction avec la PAP (Preker *et al.*, 1995). La fonction

primaire de Fip1 semble donc être de positionner la PAP dans l'environnement immédiat du site de coupure. Pour sa part, Psf2 participe à l'assemblage de la machinerie de clivage et de polyadénylation en établissant des interactions avec plusieurs sous-unités (Ohnacker *et al.*, 2000). Le complexe CFIA lie, de par sa sous-unité Rna15, des séquences riches en U situées en aval du site de coupure (Baejen *et al.*, 2014, Mayer *et al.*, 2012), un positionnement également observé chez les mammifères (Martin *et al.*, 2012). Finalement, CFIB s'associe à l'élément EE (Kessler *et al.*, 1997).

Beaucoup moins d'informations sont actuellement disponibles sur la composition de la machinerie de clivage et de polyadénylation chez *S. pombe*. Une certaine conservation semble toutefois se dégager des études protéomiques effectuées (Roguev *et al.*, 2004, Vanoosthuysen *et al.*, 2014). Cette conservation est également observable fonctionnellement. Ainsi, une version mutante de Pfs2 et de Rna14 engendrent des défauts de maturation 3' et de terminaison de la transcription (Sonkar *et al.*, 2016, Wang *et al.*, 2005). L'analyse détaillée des sous-complexes formant la machinerie de clivage et de polyadénylation reste cependant à être établie de même que le réseau d'interaction protéine-ARN.

5. Les protéines liant les queues poly(A) (PABP)

Une fois le clivage endonucléolytique du pré-ARNm effectué, une queue poly(A) est ajoutée à l'extrémité 3' du produit de clivage en amont. Les queues poly(A) sont présentes à l'extrémité 3' de la quasi-totalité des ARNm contribuant à leur stabilité et leur traduction (Chen et Shyu, 2011, Preiss et Hentze, 1998). L'influence sur ces processus ne dépend toutefois pas de la seule présence de la queue poly(A), mais requiert la participation de protéines qui viennent s'y lier avec une haute affinité et spécificité. Ces PABPs sont généralement conservées chez les eucaryotes, alors qu'elles sont absentes chez les procaryotes. Ces protéines sont divisées en deux grandes familles sur la base de leur localisation nucléaire (PABPN) ou cytoplasmique (PABPC). Parmi les différentes protéines liant les queues poly(A), PABPN1 et PABPC1 sont les représentants nucléaires et cytoplasmiques les plus étudiés et les mieux caractérisés (Kuhn et Wahle, 2004, Mangus *et*

al., 2003). Cela dit, à des fins de limitation et de pertinence, seul PABPN1 sera subséquemment abordée.

5.1. Les domaines protéiques de PABPN1

PABPN1 est largement conservée chez les eucaryotes à l'exception de chez *S. cerevisiae* (Wigington *et al.*, 2014, Winstall *et al.*, 2000). PABPN1 possède de nombreux domaines protéiques fonctionnels (Nemeth *et al.*, 1995) (Figure 2). PABPN1 comprend, en outre, un domaine N-terminal acide caractérisé par la présence d'une suite de dix alanines suivant la méthionine initiatrice. Cette particularité est importante dans la mesure où une extension de cette séquence d'alanine est retrouvée chez des individus atteints de la dystrophie musculaire oculopharyngée (OPMD : *oculopharyngeal muscular dystrophy*) (Brais *et al.*, 1998). Le domaine N-terminale comprend également un motif structural de type *coiled-coil* essentiel pour la stimulation de l'activité de polymérisation de la poly(A) polymérase (Kerwitz *et al.*, 2003, Kuhn *et al.*, 2003). La liaison spécifique et de haute affinité à des séquences homopolymériques d'adénosines s'effectue, quant à elle, par le biais d'un RRM unique positionné entre les domaines N- et C-terminaux (Kuhn *et al.*, 2003). Le domaine C-terminale est, pour sa part, basique et ne contient aucun motif structural apparent, mais est néanmoins fonctionnellement important dans la mesure où il contribue à la fois à la liaison à l'ARN (Kuhn *et al.*, 2003) et à la poly(A) polymérase (Kerwitz *et al.*, 2003) en plus d'influencer la capacité d'oligomérisation de molécules de PABPN1 (Fan *et al.*, 2001, Kuhn *et al.*, 2003). C'est également au niveau de ce domaine que se situe le signal de localisation nucléaire (NLS : *nuclear localization signal*) (Calado *et al.*, 2000, Mallet et Bachand, 2013). Finalement, le domaine C-terminal est sujet à l'activité de multiples protéines arginine méthyle-transférases de type I (Fronz *et al.*, 2008, Smith *et al.*, 1999) ce qui a pour effet d'affaiblir l'interaction entre PABPN1 et *transportin* (karyophérine β 2/Kap β 2), son récepteur d'import nucléaire (Fronz *et al.*, 2011). Cela dit, il appert toutefois que le mécanisme d'import nucléaire de PABPN1 est largement indépendamment de Kap β 2 (Mallet et Bachand, 2013). L'importance fonctionnelle de la di-méthylation asymétrique des arginines sur les propriétés de PABPN1 est donc très énigmatique surtout considérant que

ces modifications n'affectent en rien sa capacité à lier un ARN polyadénylé (Kuhn *et al.*, 2003).

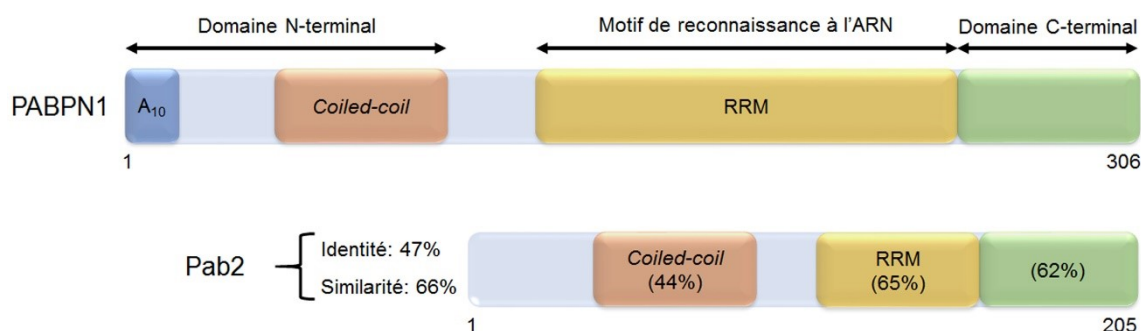


Figure 2. Schéma des domaines protéiques présents chez PABPN1 et Pab2.

Quatre domaines peuvent être discernés chez PABPN1 : une suite de 10 alanines positionnée à l'extrémité amino(N)-terminale, un domaine *coiled-coil*, un domaine de reconnaissance à l'ARN (RRM) ainsi qu'un domaine carboxy/(C)-terminale. Pab2 est l'orthologue présumé de PABPN1 chez *S. pombe*. Pab2 diffère de PABPN1 en regard de sa taille, mais également par l'absence de la suite d'alanines en N-terminal. Par contre, les trois autres domaines sont relativement bien conservés conférant au final un pourcentage d'identité et de similarité en respect de PABPN1 de 47 et 66% respectivement. Les domaines protéiques sont positionnés approximativement et ne sont pas à l'échelle.

5.2. Fonction nucléaire canonique de PABPN1

L'addition d'une queue poly(A) à l'extrémité 3' des ARNms s'effectue via l'activité enzymatique de poly(A) polymérase suivant le clivage du pré-ARNm naissant. Parmi les PAPs nucléaires, la PAP dite canonique est la plus étudiée et est conservée des levures aux mammifères (Lingner *et al.*, 1991, Raabe *et al.*, 1991, Wahle *et al.*, 1991). Chez ces derniers, la majorité des ARNms possèdent, à l'équilibre, une queue poly(A) d'une longueur d'environ 50-100 nucléotides (Chang *et al.*, 2014, Subtelny *et al.*, 2014) contrastant avec les quelques 250-300 nucléotides originellement caractérisés (Sheets et Wickens, 1989, Wahle *et al.*, 1991). Ces données proviennent cependant de contextes expérimentaux forts différents. Alors que la taille des queues poly(A) fut initialement caractérisée à partir de réactions de polyadénylation *in vitro* utilisant soit un extrait nucléaire ou des facteurs protéiques purifiés

et un substrat d'ARN synthétique, l'actuelle mesure provient plutôt du séquençage massif d'ARNs polyadénylés isolés de lignées cellulaires. Considérant l'influence de diverses déadénylases cytoplasmiques sur la taille des queues poly(A) des ARNms (Yamashita *et al.*, 2005), la divergence observée provient vraisemblablement de l'absence de facteurs protéiques modulant négativement l'étendue des queues poly(A) lors des essais de polyadénylation *in vitro*. La taille moyenne des queues poly(A) varie également d'organisme en organisme, étant d'environ 30 nucléotides chez les levures *S. pombe* et *S. cerevisiae* (Subtelny *et al.*, 2014).

In vitro, le processus de polyadénylation ne requiert que trois composantes: la PAP, le complexe CPSF ainsi que PABPN1. Lorsque combiné, une réaction de polyadénylation efficace s'enclenche menant à l'obtention de queues poly(A) d'une longueur moyenne d'environ 250 nucléotides (Bienroth *et al.*, 1993, Wahle *et al.*, 1991). La présence conjointe de ces trois composantes est requise dans la mesure où la PAP est hautement inefficace et peu spécifique. Seule, PAP possède une faible affinité pour ses substrats d'ARN, ce qui résulte en une fréquence d'association/dissociation au substrat élevée engendrant une polyadénylation lente (Wahle, 1991). Cette affinité est grandement augmentée par la présence du complexe CPSF et de PABPN1 qui, par des interactions directes, stabilisent la liaison entre la PAP et l'ARN (Kerwitz *et al.*, 2003, Murthy et Manley, 1995). Ces interactions modulent l'activité de la PAP faisant passer la réaction de polyadénylation d'un mode lent à un mode rapide (Bienroth *et al.*, 1993). La spécificité du processus de polyadénylation est également favorisée dans la mesure où le complexe CPSF recrute la PAP au niveau d'un substrat d'ARN pourvu d'un motif hexamérique AAUAAA (Keller *et al.*, 1991) tandis que PABPN1 fait de même en présence d'un substrat doté d'une répétition de quelques adénosines à son extrémité 3' (Wahle *et al.*, 1991). Cette stimulation synergique de la PAP est toutefois transitoire car passé ≈ 250 nucléotides, l'activité de la PAP redevient lente (Wahle, 1995). La synthèse des queues poly(A) peut donc être considérée comme un processus biphasique dynamiquement régulé. Cette transition de phase implique un remodelage des interactions entre le complexe CPSF, PABPN1 et la PAP (Figure 3). Durant la phase initiale de synthèse, un contact continu est présent entre ces trois facteurs. L'établissement de ce réseau d'interaction a lieu dès l'ajout par la PAP de 11-14 adénosines,

une longueur suffisante pour la liaison d'une première molécule de PABPN1 (Meyer *et al.*, 2002). Selon les données actuelles, le revêtement de la queue poly(A) en élongation par des molécules de PABPN1 mène à la formation d'une particule sphérique permettant à l'ARN d'adopter un repliement faisant en sorte que le complexe CPSF, PABPN1 et la PAP soit en communication continue (Keller *et al.*, 2000). Cependant, passé ≈ 250 nucléotides, cette particule sphérique ne permet plus d'accommoder des molécules de PABPN1 brisant ainsi l'interaction synergique entre la PAP, le complexe CPSF et PABPN1. Se faisant, toute élongation subséquente par la PAP, médiée uniquement par PABPN1, devient lente ce qui correspond à la deuxième phase du processus de polyadénylation (Kuhn *et al.*, 2009). PABPN1 peut donc être considérée comme un déterminant critique régissant la taille des queues poly(A) au noyau. La/les raison(s) derrière(s) l'accommodation d'un nombre défini de molécules de PABPN1 au sein de la structure sphérique est/sont toutefois incomprise(s).

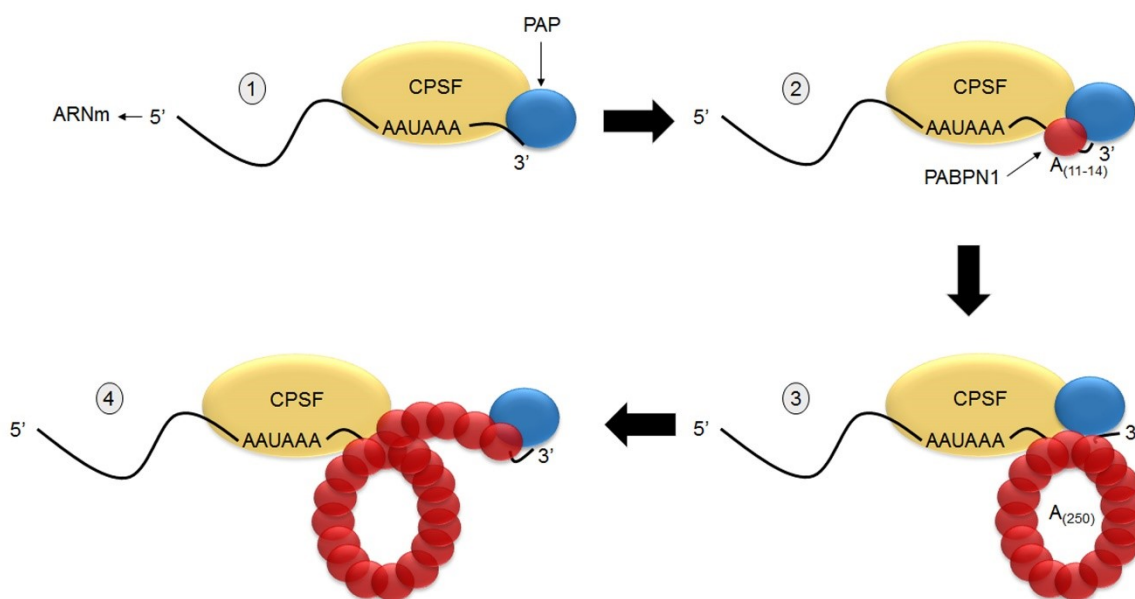


Figure 3. Schéma du mécanisme de polyadénylation PABPN1-dépendant.

Le recrutement de la poly(A) polymérase (PAP) à l'extrémité 3' d'un ARN clivé est favorisé par l'interaction entre celle-ci et le complexe CPSF (1). Cette interaction permet à PAP de synthétiser une courte queue poly(A) permettant le recrutement d'une première molécule de PABPN1 (2). La présence de PABPN1 mène à un accroissement de la cinétique de synthèse de la queue poly(A). Cette influence positive est toutefois abrogée lorsque la queue poly(A) atteint ≈ 250 nucléotides (3). À ce moment, l'interaction de la PAP avec le complexe CPSF et PABPN1 est déstabilisée menant à l'interaction de la PAP avec uniquement PABPN1 (4). Les schémas ne sont pas à l'échelle.

5.3. Pab2: l'orthologue putatif de PABPN1 chez *S. pombe*

Chez *S. pombe*, Pab2 est l'orthologue présumée de PABPN1 (Perreault *et al.*, 2007). Pab2 partage 47% d'identité et 66% de similarité au niveau de sa séquence en acide aminé et possède des domaines protéiques hautement similaires (Figure 2). Pab2 est cependant passablement plus petite que sa contrepartie humaine conséquence de l'absence d'un domaine N-terminale acide. De manière similaire à PABPN1, Pab2 est di-méthylée asymétriquement et possède un NLS au niveau de son domaine C-terminal riche en arginine. La méthylation de Pab2 n'est toutefois pas requise pour sa localisation nucléaire laquelle requiert la participation de Kap104, l'orthologue chez *S. pombe* de la Kap β 2 chez l'humain (Mallet et Bachand, 2013). Tout comme PABPN1, Pab2 possède une spécificité de liaison pour des ARNs polyadénylés et cette liaison est indépendante de son statut de méthylation. En fait, la seule fonction connue à ce jour pour cette modification post-traductionnelle est dans la limitation de l'oligomérisation de molécules de Pab2 (Perreault *et al.*, 2007). Curieusement, Pab2 ne constitue pas une protéine essentielle. Son absence n'est toutefois pas sans conséquence alors que des cellules *pab2* Δ affichent un retard de croissance à basse température (Perreault *et al.*, 2007). L'absence de Pab2 engendre également l'expression d'ARNs hyperadénylés indiquant une fonction pour Pab2 dans le métabolisme des queues poly(A) (Perreault *et al.*, 2007). La détection de ces ARNs hyperadénylés indique toutefois que, contrairement à PABPN1, Pab2 n'est nécessaire ni au recrutement de la PAP au substrat d'ARN ni à l'établissement d'une réaction de polyadénylation processive. Une fonction de Pab2 dans le contrôle de la taille des queues poly(A), telle que définie pour PABPN1, est également peu probable compte tenu que le phénotype d'hyperadénylation est limité à une sous-population d'ARNs. Ainsi, en respect des connaissances actuelles, le rôle de Pab2 au niveau du métabolisme des queues poly(A) est à être éclairci.

6. La terminaison de la transcription

La terminaison de la transcription est le processus par lequel l'ARNP se dissocie de la matrice d'ADN à la fin d'un événement de transcription. Fonctionnellement, la terminaison de la transcription est nécessaire à la partition du génome en définissant la frontière terminale

des unités géniques. Cependant, son importance va bien au-delà de la simple délimitation de l'une des deux extrémités d'un gène. La terminaison de la transcription peut également servir à réguler l'expression génique (voir la section 6.2). Malgré les contributions importantes de la terminaison transcriptionnelle au niveau cellulaire, il est pour le moins surprenant de constater que les déterminants moléculaires menant au démantèlement efficace d'un complexe d'ARNPII en élongation sont encore à ce jour largement méconnus.

6.1. La terminaison de la transcription des gènes codants

Chez les eucaryotes, l'extrémité 3' d'un ARNm est définie par la position où le clivage du pré-ARNm naissant et l'ajout d'une queue poly(A) ont lieu. Ces étapes de maturation du transcrit naissant sont intimement liées à la reconnaissance de signaux nucléotidiques spécifiques, le PAS, présents le long de son 3' UTR par un complexe multiprotéique. La terminaison de la transcription, quant à elle, s'effectue aléatoirement en aval du PAS et implique la contribution des mêmes signaux nucléotidiques et facteurs protéiques nécessaires à la maturation 3' du pré-ARNm. La maturation 3' et la terminaison transcriptionnelle sont ainsi fonctionnellement liées (Birse *et al.*, 1998). Or, deux hypothèses ont émergées au fil du temps pour expliquer l'interrelation de ces deux processus.

6.1.1. La terminaison de la transcription de type *torpedo*

Le modèle de terminaison de type *torpedo* propose que le clivage endonucléolytique du pré-ARNm naissant par la machinerie de clivage et de polyadénylation, génère un point d'entrée pour une exoribonucléase 5'→3', laquelle dégrade l'extrémité 5' monophosphate non coiffée du produit de clivage en aval jusqu'à atteindre l'ARNPII menant ainsi à la dissociation du complexe d'élongation (Connelly et Manley, 1988, Whitelaw et Proudfoot, 1986). Le modèle de terminaison *torpedo* est notamment supporté par des études impliquant l'exoribonucléase 5'→3' Rat1 de *S. cerevisiae* et son orthologue putatif chez les humains, Xrn2, dans la terminaison transcriptionnelle (Fong *et al.*, 2015, Kim *et al.*, 2004, West *et al.*, 2004). Dans les deux cas, l'inactivation/déplétion de ces exoribonucléases mène à

l'augmentation de la densité en ARNP_{II} en aval des PAS synonyme d'un défaut dans la terminaison de la transcription (Figure 4).

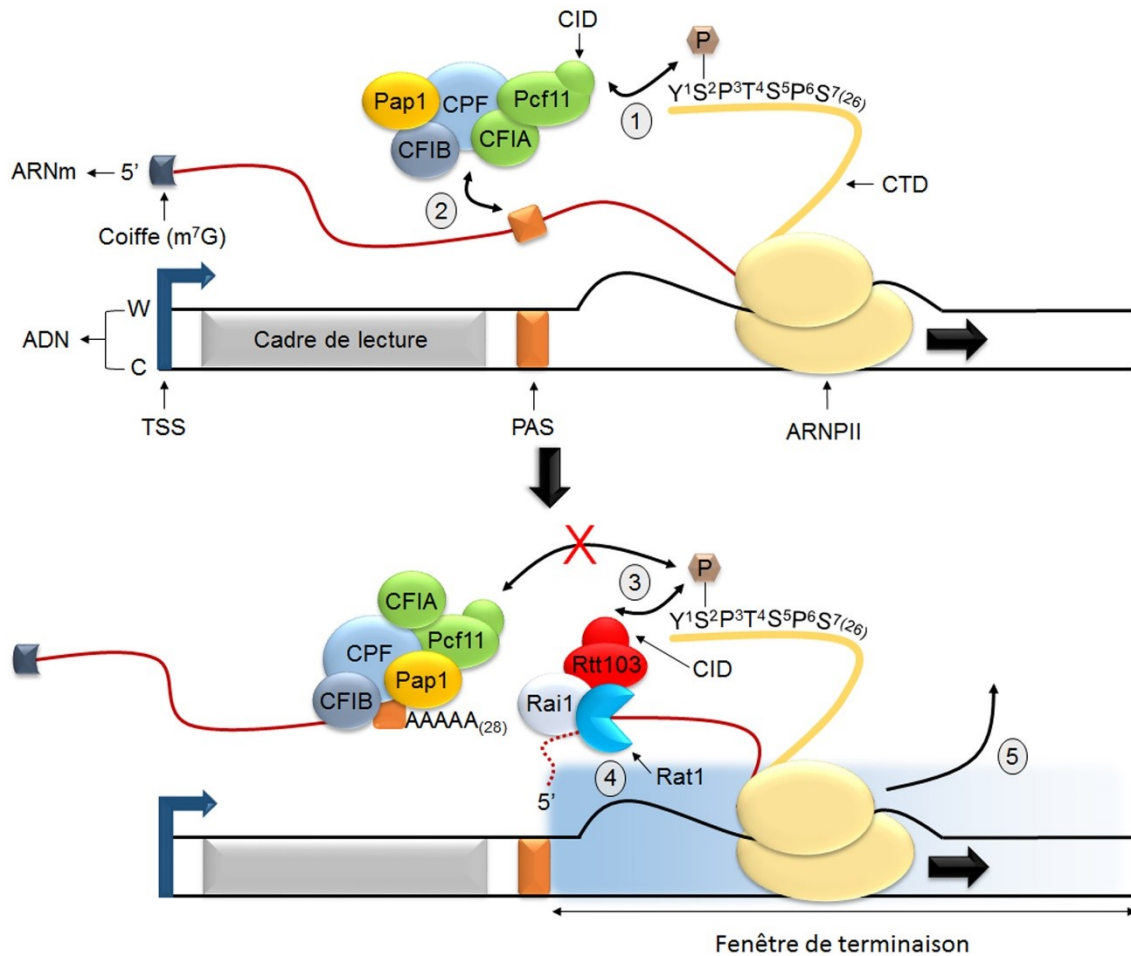


Figure 4. Schéma de la terminaison transcriptionnelle de type *torpedo*.

La terminaison transcriptionnelle de type *torpedo* implique la reconnaissance d'un domaine CTD phosphorylé en S². Cette modification est reconnue par le domaine CID de Pcf11, une sous-unité du complexe de clivage et de polyadénylation CFIA (1). Cette interaction aide au recrutement de l'ensemble des sous-complexes formant la machinerie de clivage et de polyadénylation (CF1B, CPF et Pap1) et de Rtt103. La machinerie de clivage et de polyadénylation est subséquentement transférée au niveau de l'ARN naissant suite à l'émergence du PAS (2). Au site de polyadénylation, le pré-ARNm est clivé. Le clivage endonucléolytique permet l'ajout d'une queue poly(A) au fragment en amont lequel deviendra l'ARNm ainsi que l'exposition d'une extrémité d'ARN 5' libre au niveau du fragment en aval. Ce dernier est un substrat pour l'exoribonucléase 5'→3' Rat1. Rat1 forme un complexe avec Rai1 et Rtt103. Ce dernier favorise le recrutement de Rat1 dans l'environnement du PAS par l'établissement d'une interaction entre son domaine CID et le domaine CTD phosphorylé en S² (3). Rat1 favorise la terminaison de la transcription via la

dégradation du fragment d'ARN encore attaché à l'ARNPII ce qui a pour effet de déstabiliser la bulle transcriptionnelle du complexe en élongation d'où l'analogie à la torpille (*torpedo*) (4-5). La terminaison transcriptionnelle de type *torpedo*, s'effectue aléatoirement le long d'une «fenêtre de terminaison» laquelle est influencée par les cinétiques de dégradation et de polymérisation de Rat1 et de l'ARNPII respectivement. Le schéma n'est pas à l'échelle.

6.1.2. La terminaison de la transcription de type allostérique

Alternativement, le modèle de terminaison transcriptionnelle allostérique, aussi connu comme le modèle anti-terminateur ou encore le modèle du changement de conformation, propose que la transcription d'un PAS engendre un changement structural au sein du complexe d'élongation causé par la dissociation de facteurs d'élongation et/ou l'association de facteurs de terminaison menant ultimement à une perte du potentiel d'élongation de l'ARNPII et conséquemment la terminaison de la transcription (Kim *et al.*, 2004, Logan *et al.*, 1987, Zhang *et al.*, 2015). Cette indépendance entre le processus de terminaison transcriptionnelle et l'évènement de clivage fut initialement supportée par la visualisation par microscopie électronique d'évènements de transcription chez des ovocytes de *Xenopus* lesquels ont révélé que le clivage du pré-ARNm naissant n'est pas un prérequis à la terminaison de la transcription (Osheim *et al.*, 1999). Abondant dans ce sens, une étude subséquente chez l'organisme modèle *Drosophila melanogaster* a révélé que pour la majorité des gènes analysés le transcrit synthétisé est relâché avant l'évènement de clivage (Osheim *et al.*, 2002). L'intérêt de ce modèle réside également dans le fait qu'une transition au niveau des protéines associées au complexe d'élongation s'effectue dans la région 3' des gènes (Kim *et al.*, 2004). De plus, il appert que la terminaison transcriptionnelle de type allostérique peut être inhibée par l' α -amanitine, un composé qui se lie à la *trigger loop* de l'ARNPII (Kaplan *et al.*, 2008, Zhang *et al.*, 2015). Cette donnée est en accord avec l'idée qu'un changement de conformation au niveau du complexe d'élongation suite à la transcription d'un PAS entraîne la terminaison de la transcription (Figure 5).

Il est intéressant de mentionner que Rat1 n'est pas uniquement associée au modèle de terminaison *torpedo*. Une version de Rat1 dépourvue de son activité enzymatique et inapte à déloger un complexe d'élongation *in vitro*, voit cette activité être compensée par son

association avec Rtt103. Notamment cette restauration de la capacité à induire la terminaison de la transcription dépend du domaine CTD de Rtt103. La liaison au domaine CTD par le biais de Rtt103 est ainsi nécessaire afin d'établir un positionnement adéquat de Rat1 avec l'ARNPII afin de la déloger de la matrice d'ADN probablement au travers de l'induction d'un changement allostérique (Pearson et Moore, 2013).

Pcf11 est un autre facteur protéique pourvu d'une capacité intrinsèque à terminer la transcription. La déplétion de Pcf11 mène à des défauts de clivage et de polyadénylation ainsi que de terminaison de la transcription (Sadowski *et al.*, 2003). Cela dit, une version mutante de Pcf11 compétente dans sa fonction de clivage et de polyadénylation engendre également des défauts de terminaison de la transcription suggérant que le rôle de Pcf11 dans la terminaison transcriptionnelle n'est pas uniquement de créer un point d'entrée pour Rat1/Xrn2. Or, il est difficile de séparer les fonctions moléculaires attribuées à Pcf11 de celles associées à Rat1 compte tenu que le recrutement de Rat1 et Pcf11 à l'extrémité 3' des gènes est étroitement lié (Lunde et al., 2010, Luo et al., 2006). Cela dit, chez *S. cerevisiae* et *Drosophila melanogaster*, Pcf11 peut, de manière indépendante, déstabiliser un complexe d'élongation *in vitro* par le truchement d'interactions simultanées avec le domaine CTD et l'ARN (Zhang *et al.*, 2005, Zhang et Gilmour, 2006).

Outre le domaine CTD, pas moins de 5 domaines conservés au cours de l'évolution et exposés à la surface du complexe de l'ARNPII existent lesquels pourraient interagir avec des facteurs aux propriétés favorisant la terminaison de la transcription (Garcia-Lopez et Navarro, 2011). À cet égard, des mutations dans l'hétérodimère Rpb3/Rpb11 engendrent des défauts de terminaison transcriptionnelle suggérant qu'un signal de terminaison de la transcription est transmis par l'entremise de ces sous-unités (Steinmetz *et al.*, 2006).

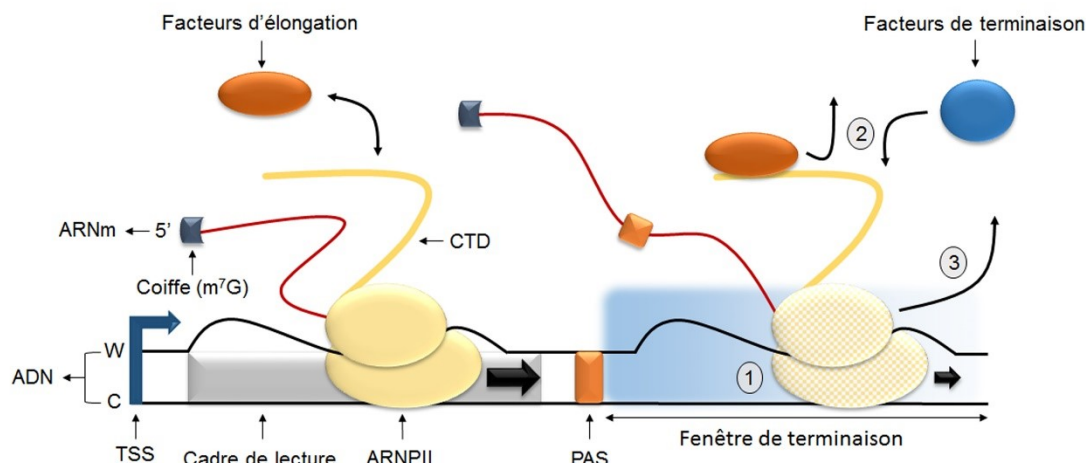


Figure 5. Schéma de la terminaison transcriptionnelle de type allostérique.

Ce mode de terminaison exclut une contribution du clivage endonucléolytique du pré-ARNm dans l'éviction de l'ARNPII de la matrice d'ADN. Dans ce mode de terminaison transcriptionnelle, le relâchement de l'ARNPII de l'ADN dépend d'un changement allostérique au niveau du complexe d'élongation (illustré par ici par une ARNPII à damier) induit par la transcription du PAS et/ou un échange entre des facteurs d'élongation et de terminaison de la transcription en aval du PAS (1-2). Ces événements modulent négativement la cinétique d'élongation de l'ARNPII favorisant ainsi son éviction à l'intérieur d'une «fenêtre de terminaison» (3). Le schéma n'est pas à l'échelle.

6.1.3. Unification des modèles de terminaison transcriptionnelle

Les modèles de terminaison de type *torpedo* et allostérique ne sont pas mutuellement exclusifs et de nombreuses indications suggèrent un rapprochement entre les deux (Luo *et al.*, 2006, Zhang *et al.*, 2015). Notamment, des données *in vitro* indiquent que le processus de terminaison de la transcription s'effectue sans la présence de clivage au niveau de du pré-ARNm naissant. Toutefois, malgré que le clivage et la digestion subséquente de l'ARN ne soient pas absolument nécessaires à la terminaison de la transcription *in vitro*, l'inhibition de l'activité 5'→3' exoribonucléase de Xrn2/Rat1 entraîne un délai significatif dans l'éviction du complexe d'élongation de la matrice d'ADN suggérant un effet combinatoire des modèles allostérique et *torpedo* pour l'obtention d'une terminaison de la transcription optimale (Zhang *et al.*, 2015). De manière complémentaire, l'inhibition de Xrn2 *in vivo* n'abolit pas totalement la terminaison transcriptionnelle, mais la retarde (Fong *et al.*, 2015). De plus, chez *S. cerevisiae*, la déplétion de la sous-unité catalytique Ysh1 voit l'emplacement où l'ARNPII

est normalement relâchée être allongé en aval suggérant que la terminaison de la transcription s'effectue de manière sous-optimale en absence d'un clivage au niveau de pré-ARNm (Schaughency *et al.*, 2014).

L'efficacité avec laquelle la terminaison de la transcription s'accomplit est de plus étroitement liée à la nature du PAS transcrit. Ainsi, un pré-ARNm pourvu d'un PAS canonique/fort est plus enclin à être clivé et conséquemment être sujet à l'action concertée des mécanismes de terminaison transcriptionnelle de type *torpedo* et allostérique qu'un pré-ARNm doté d'un PAS non-canonique/faible lequel verrait la terminaison de sa transcription dépendre principalement d'un mécanisme de type allostérique (Chao *et al.*, 1999, Orozco *et al.*, 2002, Zhang *et al.*, 2015) (Figure 6).

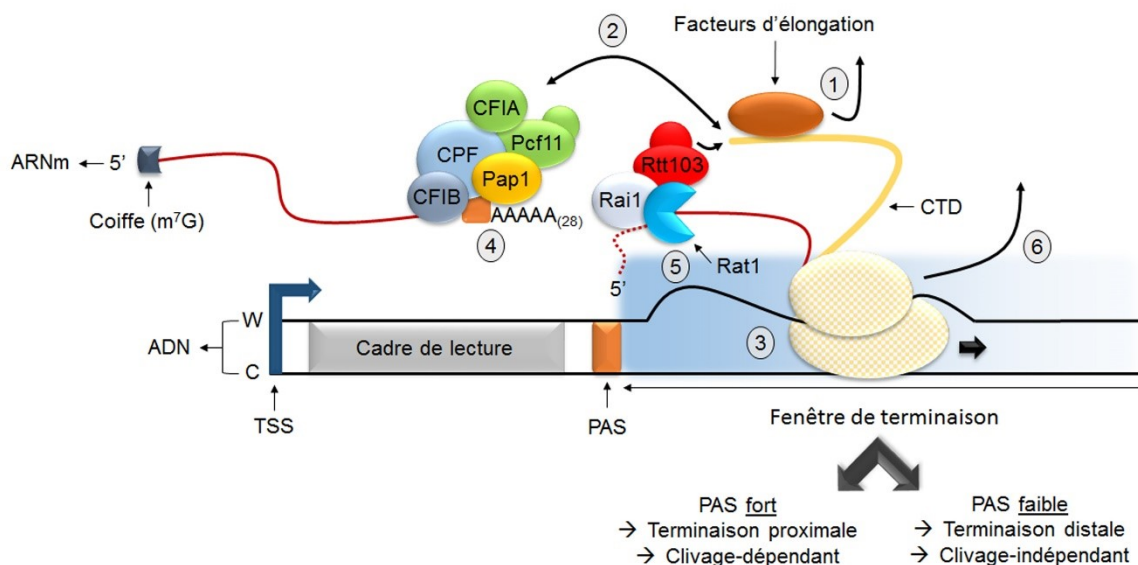


Figure 6. Schéma de l'unification des mécanismes de terminaison transcriptionnelle *torpedo* et allostérique.

Dans ce modèle unifié, la transcription d'un PAS est liée à un changement au niveau des facteurs associés au complexe d'élongation (1-2). La transcription du PAS induit également un changement de conformation au niveau de l'ARNPII modulant négativement sa capacité d'élongation (3). L'emplacement où l'ARNPII est évincé de l'ADN («fenêtre de terminaison») est fortement influencé par la nature du PAS. Ce dernier est ainsi un élément central du modèle de terminaison transcriptionnelle unifiée. En présence d'un PAS consensus (fort), le pré-ARNm subit un clivage endonucléolytique permettant l'activité de Rat1 (4-5). La combinaison de la décélération de la cinétique de transcription et de l'activité *torpedo* de Rat1 favorise ainsi une terminaison transcriptionnelle dite proximale (6). En contrepartie, un PAS dégénéré (faible) ne favorise pas le clivage endonucléolytique du pré-ARNm abrogeant

l'activité de terminaison de la transcription de Rat1. Dans une telle circonstance, la «fenêtre de terminaison» transcriptionnelle s'étend (terminaison distale) (6). Le schéma n'est pas à l'échelle.

6.1.4. L'association entre la pause transcriptionnelle et la terminaison de la transcription.

Un déterminant clé de l'efficacité avec laquelle la terminaison transcriptionnelle a lieu est sans contredit lié à la modulation de la vitesse d'élongation de l'ARNPII en aval du PAS. En défavorisant la cinétique d'élongation de l'ARNPII, la probabilité qu'une interaction se produise entre un facteur engendrant la terminaison et l'ARNPII augmente. Ce concept est notamment important pour l'activité de terminaison de Rat1/Xrn2. Ainsi, la «fenêtre de terminaison» Rat1/Xrn2-dépendante est raccourcie lorsque la transcription est effectuée par une version «lente» de l'ARNPII alors qu'elle s'allonge en présence d'une ARNPII «rapide» (Fong *et al.*, 2015). Cette compétition cinétique est notamment influencée par la propension qu'à l'ARNPII à pauser en aval d'un PAS. Cette caractéristique se reflète par une accumulation d'ARNPII à l'extrémité 3' des gènes (Anamika *et al.*, 2012, Coudreuse *et al.*, 2010). Cette accumulation est notamment une conséquence de la transcription du PAS ainsi que de l'influence de la machinerie de clivage et de polyadénylation (Nag *et al.*, 2007, Nojima *et al.*, 2015, Orozco *et al.*, 2002). De plus, la présence de certaines séquences d'ADN situées en aval du PAS favorise l'état de pause au niveau de l'ARNPII (Birse *et al.*, 1997, Gromak *et al.*, 2006). Ces séquences d'ADN peuvent influencer le processus transcriptionnel directement en créant un hybride ARN:ADN instable au niveau de la bulle transcriptionnelle de l'ARNPII, une condition favorable à l'établissement d'une pause au niveau de la transcription (Kulish et Struhl, 2001, von Hippel, 1998). Une corrélation existe également entre le contexte chromatinien, à savoir la présence de nucléosomes, et l'accumulation d'ARNPII à l'extrémité 3' des gènes (Grosso *et al.*, 2012). Dans tous ces cas, l'objectif est de limiter le potentiel d'élongation de l'ARNPII afin d'optimiser les processus de maturation 3' et de terminaison de la transcription.

6.2. La terminaison transcriptionnelle d'ARN non-codants chez la levure

Outre les ARNm, l'ARNPII transcrit un nombre imposant d'ARNncs chez la levure. Certains de ces ARNncs exerce des fonctions importantes au niveau cellulaire, tels que les *snoRNAs* ou encore les *snRNAs*, alors que la majorité restante représente le produit d'une transcription généralisée du génome dont les fonctions biologiques font l'objet d'intenses spéculations.

6.2.1. La terminaison transcriptionnelle NNS-dépendante

Chez *S. cerevisiae*, le complexe NNS, composé des protéines Nab3 (*nuclear polyadenylated RNA-binding protein 3*) (Wilson *et al.*, 1994), Nrd1 (*nuclear pre-mRNA downregulation protein 1*) (Steinmetz et Brow, 1996) et Sen1 (*splicing endonuclease 1*) (Winey et Culbertson, 1988), est important pour la terminaison de la transcription de nombreux ARNncs tels que les *snoRNAs*, les *snRNAs* ainsi qu'un nombre important d'ARNs cryptiques issus de la transcription généralisée du génome (Arigo *et al.*, 2006, Marquardt *et al.*, 2011, Rasmussen et Culbertson, 1998, Steinmetz *et al.*, 2001, Thiebaut *et al.*, 2006). Pour promouvoir un tel évènement, le complexe NNS nécessite l'établissement d'interactions avec l'ARN naissant et le domaine CTD de l'ARNPII. La liaison à l'ARN s'effectue par le biais de Nab3 et Nrd1 lesquels possèdent chacun un RRM permettant de reconnaître des séquences d'ARNs spécifiques (UCUUG et GUAA/G pour Nab3 et Nrd1 respectivement) (Carroll *et al.*, 2007, Carroll *et al.*, 2004, Creamer *et al.*, 2011, Wlotzka *et al.*, 2011). Ces motifs sont fréquemment regroupés et coopèrent avec des séquences avoisinantes afin d'accroître l'efficacité de la terminaison de la transcription (Bacikova *et al.*, 2014, Porrua *et al.*, 2012). L'efficacité est également grandement influencée par la formation d'un complexe stable entre Nab3 et Nrd1 lequel permet d'augmenter significativement l'affinité aux motifs d'ARN (Carroll *et al.*, 2007, Conrad *et al.*, 2000, Hobor *et al.*, 2011). L'interaction avec l'ARNPII s'effectue, quant à elle, par l'entremise de Nrd1 via son domaine CID (Steinmetz et Brow, 1996, Yuryev *et al.*, 1996) et requiert un domaine CTD phosphorylé en S⁵ (Kubicek *et al.*, 2012, Vasiljeva *et al.*, 2008). Cette modification post-traductionnelle du CTD étant présente tôt durant le cycle transcriptionnelle (Komarnitsky *et al.*, 2000), l'interaction

$CID_{Nrd1p}/CTD_{ARNP_{II}}$ est considérée comme un déterminant important de l'utilisation préférentielle de cette voie de terminaison de la transcription envers les gènes de petite taille (Gudipati *et al.*, 2008, Heo *et al.*, 2013, Jenks *et al.*, 2008, Steinmetz *et al.*, 2006). L'actuelle fonction de terminaison de la transcription du complexe NNS est toutefois associée à l'activité hélicase 5'→3' ADN/ARN-dépendante de Sen1 (Martin-Tumasz et Brow, 2015). Pour en arriver à un tel résultat, Sen1 est initialement recruté au niveau de l'ARNP_{II} (Ursic *et al.*, 2004) par un mécanisme requérant la participation du facteur de clivage et de polyadénylation Pcf11 et un domaine CTD phosphorylé en S² (Ballarino *et al.*, 2005, Chinchilla *et al.*, 2012, Grzechnik *et al.*, 2015, Lenstra *et al.*, 2013, Steinmetz *et al.*, 2001, Tomson *et al.*, 2013). Pcf11 participe également à la déstabilisation de la liaison entre Nrd1 et l'ARNP_{II} (Grzechnik *et al.*, 2015). Ce faisant, Pcf11 favorise l'interaction entre Sen1 et Nab3 (Nedeja *et al.*, 2008) et conséquemment l'association de Sen1 avec l'ARN naissant. Sen1 peut dès lors se déplacer le long du transcrit naissant afin de déstabiliser l'ARNP_{II} en cours d'élongation via son activité hélicase (Martin-Tumasz et Brow, 2015, Porrua et Libri, 2013). La terminaison de la transcription par le complexe NNS est directement influencée par les cinétiques de déplacement de Sen1 et de l'ARNP_{II} et est influencée par la présence en région 3' des gènes sujets à une terminaison de la transcription NNS-dépendante d'hybrides ARN:ADN (structure *R-loop*) (Grzechnik *et al.*, 2015, Hazelbaker *et al.*, 2013, Lenstra *et al.*, 2013). Ces structures ont notamment été associées à la présence de pause transcriptionnelle (Skourti-Stathaki *et al.*, 2011). Somme toute, même si ce mode de terminaison de la transcription ne requiert pas de clivage endonucléolytique du pré-ARN, il partage néanmoins des similarités frappantes avec la terminaison de type *torpedo*. Ainsi, la terminaison NNS-dépendante nécessite: 1) la reconnaissance de modifications post-traductionnelles spécifiques du domaine CTD de l'ARNP_{II}, 2) la liaison à des motifs d'ARN définis et 3) l'éviction de l'ARNP_{II} de la matrice l'ADN selon un mécanisme d'action 5'→3' (Figure 7).

Cela dit, malgré les nombreuses études soulignant l'importance du complexe NNS dans la terminaison d'ARN_{ncs} chez *S. cerevisiae*, la conservation d'un tel mécanisme chez les eucaryotes est largement méconnue. Or, *S. pombe* possède des orthologues putatifs pour chacune des composantes du complexe NNS et Seb1, l'orthologue présumé de Nrd1 chez *S.*

pombe, participe à l'établissement d'hétérochromatine via sa liaison à des transcrits non-codants d'origine péricentromériques (Marina *et al.*, 2013) rappelant la fonction répressive de Nrd1 aux loci d'ADN ribosomale et télomérique (Vasiljeva *et al.*, 2008). De plus, l'organisation génomique des gènes de *snoRNAs* est hautement similaire entre *S. pombe* et *S. cerevisiae* alors que la majorité des *snoRNAs* sont transcrits à partir d'unités monocistroniques (Dieci *et al.*, 2009). Bien que circonstancielle, ces données laissent présager une possible conservation du mécanisme de terminaison de la transcription NNS chez *S. pombe*.

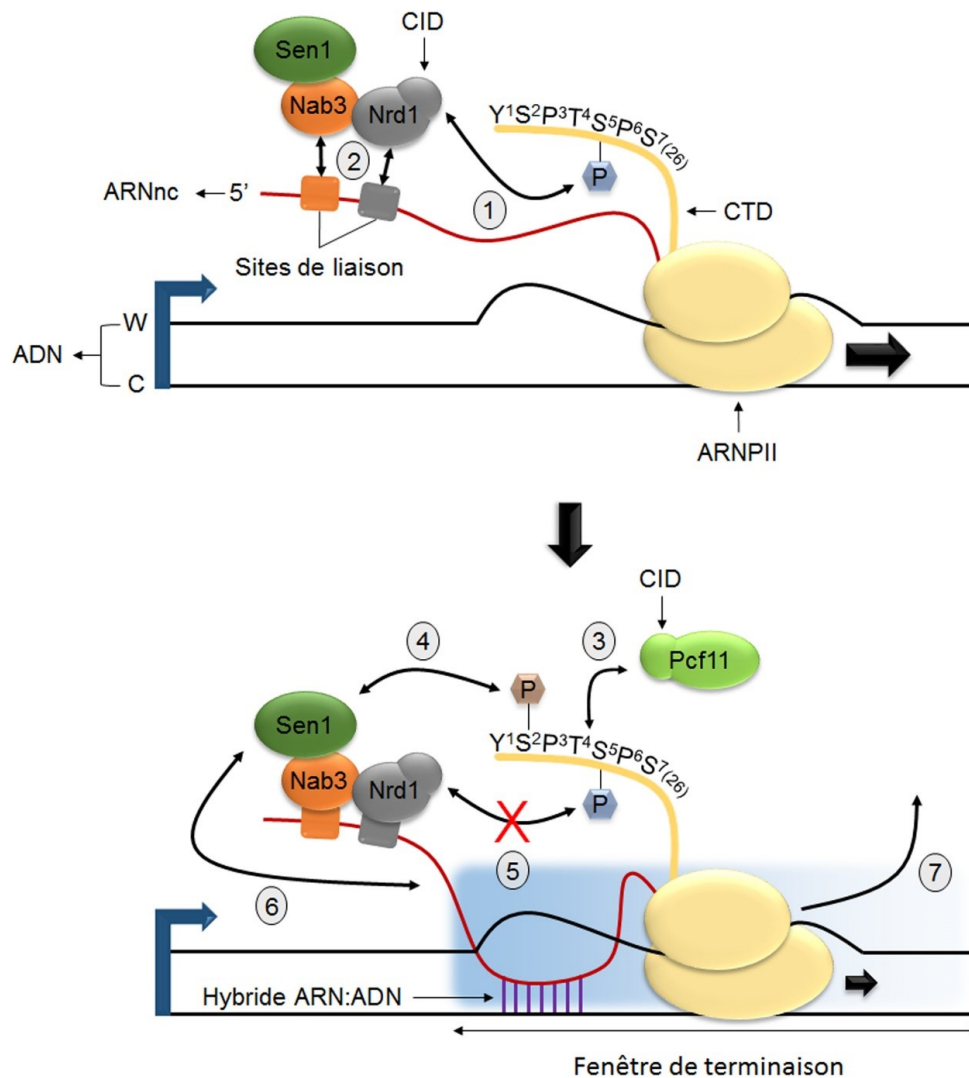


Figure 7. Schéma de la terminaison transcriptionnelle NNS-dépendante.

Chez *S. cerevisiae*, ce mécanisme de terminaison de la transcription agit principalement au niveau des gènes non-codants de petite taille. Ceux-ci sont notamment associés à une présence prépondérante d'ARNPII dotées d'un domaine CTD phosphorylé en S⁵. Cette modification du domaine CTD est reconnue par le domaine CID de Nrd1 (1). Nrd1, en complexe avec Nab3, interagit subséquemment avec l'ARN naissant via la reconnaissance de motifs d'ARN spécifiques (2). L'actuelle activité permettant l'éviction de l'ARNPII de la matrice d'ADN est toutefois associée à Sen1, une hélicase ARN-dépendante 5'→3'. Pcf11 participe à la liaison de Sen1 au domaine CTD en favorisant la phosphorylation de la S² suite à sa liaison au domaine CTD par son domaine CID (3-4). Pcf11 agit également de manière à transférer le complexe de terminaison de la transcription NNS sur l'ARN naissant en défavorisant l'interaction entre Nrd1 et le domaine CTD (5). À cette étape, Sen1 est transféré du domaine CTD à l'ARN naissant via sa capacité à interagir avec Nab3. Une fois sur l'ARN, Sen1 se déplace le long de celui-ci jusqu'à atteindre et déloger l'ARNPII (6-7). Dans ce mode de terminaison, l'éviction de l'ARNPII a lieu aléatoirement dans une «fenêtre de terminaison» laquelle est influencée par les cinétiques de déplacement de Sen1 et de l'ARNPII. Sen1 en sort gagnant dans la mesure où l'ARNPII est sujet à de la pause transcriptionnelle telle que révélée par la présence d'hybrides ARN:ADN (*R-loops*) en amont du complexe d'élongation. Le schéma n'est pas à l'échelle.

7. Les petits ARN nucléolaires (*snoRNAs*)

La modification post-transcriptionnelle des ARNs est une caractéristique commune aux domaines *Archea*, *Bacteria* et *Eukaryota*. Près de 150 altérations chimiques de nucléoside ont notamment été identifiées (Machnicka *et al.*, 2013). De ce nombre, la pseudouridylation (ψ) et la méthylation sont les deux changements les plus prévalents (Motorin et Helm, 2011, Yu et Meier, 2014). Ces modifications sont assurées par deux mécanismes enzymatiques distincts qui diffèrent selon la stratégie employée afin de reconnaître l'ARN cible. Essentiellement, le premier mécanisme dépend de la capacité intrinsèque de certaines enzymes (*stand-alone protein*) à lier spécifiquement un substrat d'ARN alors que le second requiert plutôt la participation d'un complexe enzymatique multiprotéique doté d'un ARN guide reconnaissant par complémentarité de séquence un ARN cible. Ces ARNs guides font partis de la famille des *snoRNAs*. Les *snoRNAs* se divisent en deux classes selon la présence d'éléments de séquence et de structures caractéristiques, du complexe ribonucléoprotéique formé (*snoRNP* : *small-nucleolar ribonucleoprotein*) et de la modification chimique apportée à l'ARN. Ainsi, les *snoRNPs* catalysant l'isomérisation de l'uridine en pseudouridine sont de type H/ACA (Ganot *et al.*, 1997) alors que ceux associés à la 2'-O-méthylation (addition

d'un groupement méthyl au groupement hydroxyle en position 2' de la partie glucidique d'un nucléoside) sont de type C/D (Kiss-Laszlo *et al.*, 1996). Ces modifications sont particulièrement abondantes au niveau de l'ARNr où elles se concentrent autour de domaines importants du ribosome (Decatur et Fournier, 2002) suggérant un rôle potentiel pour ces modifications dans le repliement et la stabilité de l'ARNr (Auffinger et Westhof, 1997, Davis, 1995, Kierzek *et al.*, 2014) et par conséquent le fonctionnement adéquat du ribosome. Cependant, la majorité des *snoRNAs* peuvent être supprimés individuellement du génome sans engendrer d'effets délétères importants (Piekna-Przybylska *et al.*, 2007), indiquant que l'optimisation du ribosome résulte plutôt d'une combinaison de ces modifications (Baxter-Roshek *et al.*, 2007, King *et al.*, 2003, Tollervey *et al.*, 1993, Zebarjadian *et al.*, 1999). Outre la modification de l'ARNr, certains *snoRNAs*, à la fois de type H/ACA et C/D, sont requis pour la maturation de précurseurs d'ARNr (Watkins et Bohnsack, 2012). Ces *snoRNAs* sont les seuls essentiels à la viabilité cellulaire (Piekna-Przybylska *et al.*, 2007). Chez les eucaryotes, le répertoire de substrat sujet à la pseudouridylation et/ou la 2'-O-méthylation *snoRNA*-dépendante n'est toutefois pas restreint à l'ARNr et inclut les *snRNAs* (Darzacq *et al.*, 2002, Ganot *et al.*, 1999, Ma *et al.*, 2005, Tycowski *et al.*, 1998), les *snoRNAs* ainsi que certains ARNms (Carlile *et al.*, 2014, Kishore *et al.*, 2013, Schwartz *et al.*, 2014). Au-delà de leur capacité à modifier différentes classes d'ARN, les *snoRNAs* sont également pourvus de nombreuses fonctions non-canoniques (Dupuis-Sandoval *et al.*, 2015). L'importance des *snoRNAs* est de plus soulignée par leurs associations avec plusieurs pathologies chez l'humain (McMahon *et al.*, 2015, Stepanov *et al.*, 2015).

Structurellement, les *snoRNAs* de type H/ACA adoptent un repliement secondaire caractéristique défini par la présence de deux tiges-boucles (*hairpin*) reliées entre elles par une courte région simple-brin (*hinge*). Cette dernière contient la boîte H (ANANNA) qui est l'un des deux éléments de séquences spécifiques aux *snoRNAs* de type H/ACA. L'autre élément, la boîte ACA, se positionne, quant à elle, exactement 3 nucléotides en amont de l'extrémité 3' mature du *snoRNA* (Ganot *et al.*, 1997). Chaque tige-boucle comprend également une boucle interne, surnommée la pochette de pseudouridylation, permettant le positionnement adéquat de l'uridine à modifier au niveau de l'ARN cible (Ganot *et al.*, 1997). Cette modification requiert l'activité enzymatique de Cbf5 (Lafontaine *et al.*, 1998),

une Ψ synthase, qui forme, lorsque qu'associée à Nop10 (*nucleolar protein 10*), Nhp2 (*non-histone chromosome protein 2*) et Gar1 (*glycine arginine riche protein 1*), un *snoRNP* mature fonctionnel (Yu et Meier, 2014) (Figure 8A).

Les *snoRNAs* de type C/D, pour leur part, sont caractérisés par la présence de séquences conservées dénommées boîte C (séquence consensus : RUGAUGA) et boîte D (séquence consensus : CUGA) présentes près des extrémités 5' et 3' respectivement. Ces *snoRNAs* sont également pourvus de motifs C' et D', positionnés en zone médiale du *snoRNA*, qui, chez les eucaryotes, s'apparentent aux boîtes C et D quoique moins bien conservés et souvent dégénérés (Kiss-Laszlo *et al.*, 1998). Au niveau de leur structure, les *snoRNAs* à boîtes C/D sont définis par la présence d'un motif essentiel *kink-turn*, formé de l'alignement des boîtes C et D (Watkins *et al.*, 2002), ainsi que de deux régions guides simple brin localisées immédiatement en amont des boîtes D et D' permettant la méthylation site-spécifique d'un ARN cible (Kiss-Laszlo *et al.*, 1996, Kiss-Laszlo *et al.*, 1998, Nicoloso *et al.*, 1996). Comme pour les *snoRNAs* de type H/ACA, les *snoRNAs* à boîtes C/D s'associent avec des facteurs protéiques hautement conservés afin d'obtenir un *snoRNP* fonctionnel mature. Un tel complexe comprend deux copies de la méthylase Nop1 (Tollervey *et al.*, 1993) ainsi qu'une copie des protéines Nop56, Nop58 et Snu13 qui s'assemblent de manière asymétrique autour du *snoRNA* (Cahill *et al.*, 2002, Szewczak *et al.*, 2002, Watkins *et al.*, 2000) (Figure 8B).

Qu'ils soient de type H/ACA ou C/D, l'obtention d'un *snoRNP* mature fonctionnel nécessite une succession d'évènements d'assemblage et de maturation co- et/ou post-transcriptionnels (Ballarino *et al.*, 2005, Darzacq *et al.*, 2006, Hirose *et al.*, 2003, Morlando *et al.*, 2004, Richard *et al.*, 2006, Yang *et al.*, 2005) requérant la participation de nombreuses chaperones (Bizarro *et al.*, 2014, Dez *et al.*, 2002, Fatica *et al.*, 2002, King *et al.*, 2001, Yang *et al.*, 2002, Zhao *et al.*, 2008) ainsi qu'un ciblage intranucléaire approprié (Kiss *et al.*, 2006).

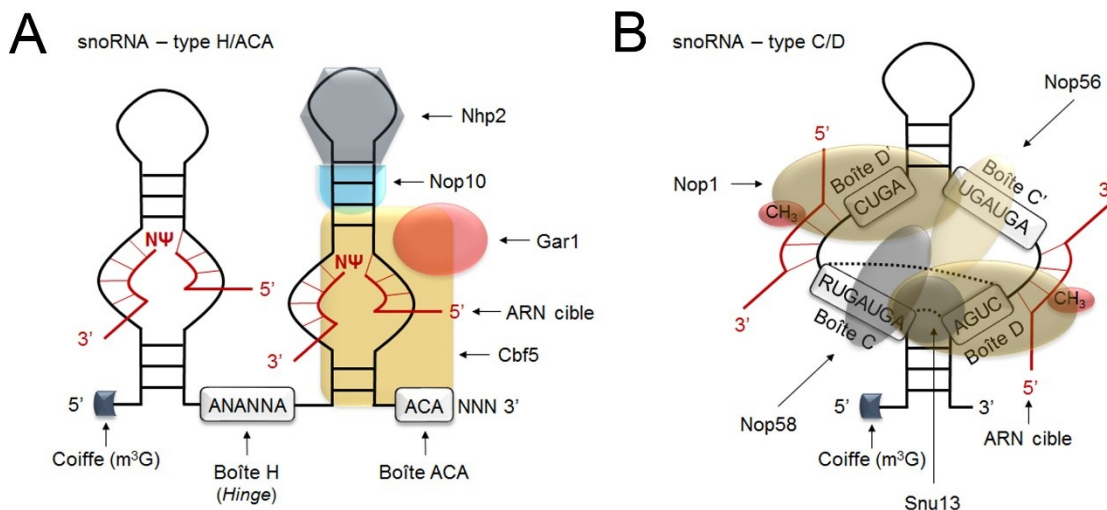


Figure 8. Schéma des différents types de *snoRNAs*.

A: Les *snoRNAs* de type H/ACA sont définis par la présence d'une structure secondaire formée de deux tige-boucles et de motifs nucléotidiques spécifiques (boîte H et ACA). Quatre protéines (Cbf5, Nop10, Nhp2, Gar1) s'agencent autour d'une structure tige-boucle afin de former un *snoRNP* mature et fonctionnel. Les *snoRNAs* de types H/ACA participent à la pseudouridylation de nombreux ARNs via l'activité pseudouridyl-synthase de Cbf5. B: Les *snoRNAs* de type C/D sont formés, quant à eux, d'une unique tige-boucle et sont pourvus d'éléments de séquences caractéristiques (boîtes C, D, C' et D'). Nop1 lie les boîtes D et D' alors que les boîtes C et C' s'associent respectivement à Nop58 et Nop56. Une quatrième protéine, Snu13, se positionne quant à elle à la jonction des boîtes C et D. Les *snoRNAs* de type C/D participent à l'ajout d'un groupement 2'-O-méthyl au niveau d'ARNs cibles. Tout comme les ARNs cibles des *snoRNAs* de type H/ACA, les ARNs ciblés par les *snoRNAs* de type C/D s'apparient à des régions simple brin présentes à même la structure tige-boucle. Nop1 est la 2'-O-méthylase. Les schémas ne sont pas à l'échelle.

8. La biogenèse des *snoRNAs*

Une caractéristique importante et distinctive des transcrits issus de la terminaison NNS-dépendante réside dans la capacité qu'ont ces transcrits à être dirigés rapidement vers l'exosome à ARN nucléaire, un complexe multiprotéique doté d'activités endo et exoribonucléase (voir la section 9), afin d'être maturés tel que lors de la formation de l'extrémité 3' des *snoRNAs* et des *snRNAs* (Allmang *et al.*, 1999, van Hoof *et al.*, 2000) ou dégradés comme dans le cas d'ARNs non-codants cryptiques (CUTs) (Davis et Ares, 2006, Wyers *et al.*, 2005). Notamment, le choix entre ces deux options semble être lié à la formation d'un complexe ribonucléoprotéique stable empêchant la progression/passage de l'exosome à

ARN (Costello *et al.*, 2011). En accord avec cette hypothèse, la déplétion des constituants d'un *snoRNP* entraîne une réduction marquée du niveau de divers *snoRNAs* matures (Henras *et al.*, 1998, Lafontaine *et al.*, 1998, Lafontaine et Tollervey, 1999, Tollervey *et al.*, 1991).

Chez *S. cerevisiae*, la génération de l'extrémité 3' mature d'un *snoRNA* est couplée à la terminaison de sa transcription (Grzechnik et Kufel, 2008). Ce couplage implique un réarrangement des interactions protéiques chez Nrd1 afin de passer d'un mode de terminaison de la transcription à un mode de maturation/dégradation. Dans les deux cas, ces interactions sont médiées par le domaine CID de Nrd1. Durant la phase de terminaison de la transcription Nrd1 établit une interaction avec le domaine CTD de l'ARNPII phosphorylé en S⁵ alors qu'en mode de maturation/dégradation Nrd1 interagit directement avec soit Mpp6 ou la sous-unité Trf4 du complexe TRAMP, deux cofacteur de l'exosome à ARN nucléaire (voir la section 10), via la présence chez ces derniers d'un motif mimant le domaine CTD de l'ARNPII (*CTD-like*) (Kim *et al.*, 2016, Tudek *et al.*, 2014). La liaison séquentielle et mutuellement exclusive entre le domaine CID de Nrd1 et le domaine CTD de l'ARNPII puis le motif *CTD-like* du complexe TRAMP ou de Mpp6 contribue ainsi à établir une coordination temporelle entre les événements de terminaison de la transcription et de maturation/dégradation dans le cas d'ARNncs (Figure 9).

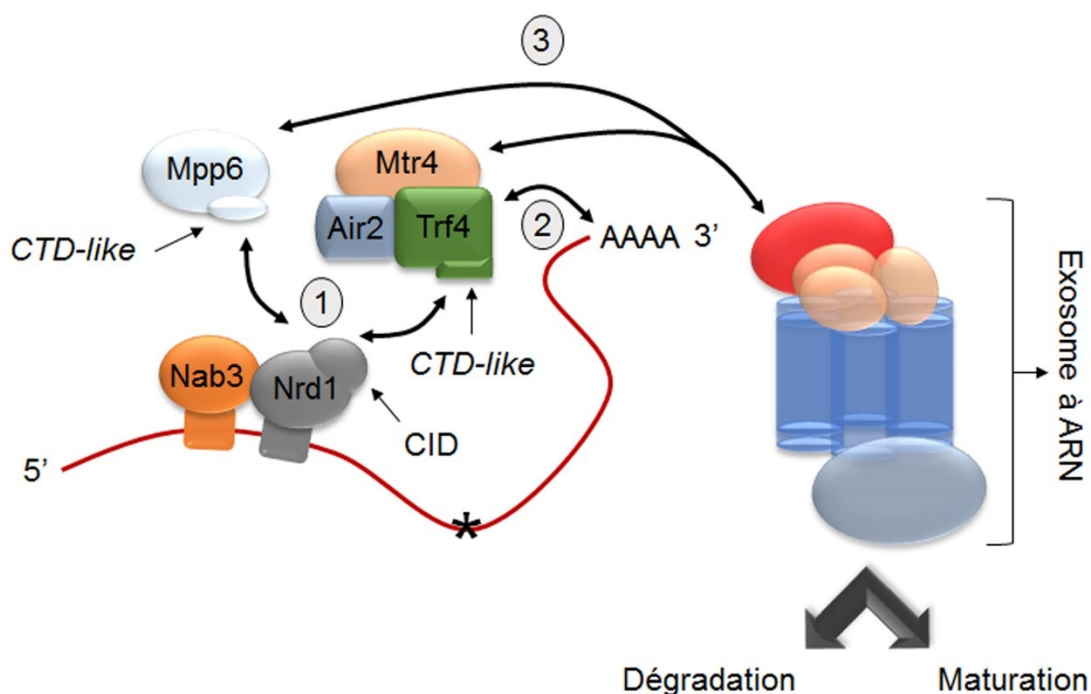


Figure 9. Schéma décrivant la connexion entre la terminaison transcriptionnelle NNS-dépendante et la maturation/dégradation d'ARNnc.

La maturation/dégradation d'un ARN dont la terminaison transcriptionnelle requiert le complexe NNS dépend de l'établissement d'une interaction entre le domaine CID de Nrd1 et d'un domaine similaire au domaine CTD de l'ARNPII (*CTD-like*) (1). Ces derniers sont retrouvés chez les cofacteurs de l'exosome à ARN Trf4 et Mpp6. Le domaine CID de Nrd1 coordonne donc temporellement les activités de terminaison transcriptionnelle et de maturation/dégradation de l'ARNnc. L'interaction avec Trf4 favorise la polyadénylation des ARNs liés par l'hétérodimère Nab3/Nrd1 et leur prise en charge par l'exosome à ARN (2-3). L'interaction avec Mpp6 favorise également l'activité de l'exosome à ARN. Dans cette situation, une activité de polyadénylation est toutefois non requise (3). * représente, dans le cas d'un *snoRNA*, l'extrémité 3' mature. La maturation/dégradation d'un *snoRNA* implique donc la prise en charge par l'exosome à ARN d'un précurseur de *snoRNA* pourvu d'une extension en 3'. Le schéma n'est pas à l'échelle.

9. L'exosome à ARN

L'exosome à ARN est un complexe multiprotéique essentiel et ubiquitaire chez les organismes eucaryotiques contribuant à la dégradation et la maturation nucléaire et cytoplasmique de nombreux types d'ARNs (Porrua et Libri, 2013). Il est bien accepté que l'activité de l'exosome à ARN s'effectue post-transcriptionnellement. Cela dit, certaines

données circonstanciées suggèrent également une connexion entre l'exosome à ARN et l'évènement transcriptionnel (Schmid et Jensen, 2013).

9.1. La structure de l'exosome à ARN

D'un point de vue architectural, l'exosome à ARN eucaryotique se présente en une unité principale (le noyau) composée de la superposition de deux structures à l'apparence d'anneau créées de l'oligomérisation de 9 polypeptides uniques. La couche inférieure est formée de l'assemblage de trois dimères (Rrp41-Rrp45, Rrp46-Rrp43 et Mtr3-Rrp42) alors que la couche supérieure, aussi connue comme le chapeau de l'exosome à ARN, est formée des protéines Csl4, Rrp4 et Rrp40 (Liu *et al.*, 2006). Une telle représentation structurale n'est toutefois pas unique au domaine *Eukarya*. L'apparence toroïdale de l'exosome à ARN eucaryotique est en fait hautement similaire à celles de complexes procaryotiques également impliqués au niveau du métabolisme d'ARNs tels que la ribonucléase PH (RNase PH) et la polynucléotide phosphorylase (PNPase) présentes au niveau du domaine *Bacteria* ou encore l'*exosome-like complex* retrouvé chez la plupart des espèces du domaine *Archaea* (Januszyk et Lima, 2011). La RNase PH présente la complexité la plus élémentaire étant composée d'un homohexamère de sous-unités à domaine PH sans présence de structure toroïde supérieure (Ishii *et al.*, 2003). La PNPase, quant à elle, se compose de trois protomères multipartites chacun composé de deux domaines *RNase PH-like* (PH-1 et PH-2) et des domaines de liaison canonique à l'ARN KH (*ribonucleoprotein K Homology*) (Nicastro *et al.*, 2015) et S1 (*ribosomal protein S1*) (Bycroft *et al.*, 1997), ces derniers se déployant spatialement de manière à générer l'architecture typique à double anneaux (Shi *et al.*, 2008). De manière similaire, l'exosome du domaine *Archaea* consiste en un anneau hexamérique formé de la trimérisation d'hétérodimères de Rrp41 et Rrp42, possédant respectivement des domaines semblables aux domaines PH-2 et PH-1 de la PNPase, lesquels sont chapeautés, dépendamment de l'espèce d'*Archaea* en présence, de trois copies de Rrp4 ou d'une combinaison de Rrp4 et Csl4, deux protéines dotés de domaines S1 et KH ou S1 respectivement (Evguenieva-Hackenberg *et al.*, 2014). De tels domaines sont également présents au niveau de l'exosome à ARN eucaryotique. En fait, les six sous-unités de son anneau inférieur partagent des similarités de séquence et de structure avec soit la composante

archéenne Rrp41 ou le domaine RNase PH-2 de la PNPase (Rrp41, Mtr3 et Rrp46) ou avec la composante archéenne Rrp42 ou le domaine RNase PH-1 de la PNPase (Rrp42, Rrp43 et Rrp45). Cette conservation de domaine est également présente au niveau des sous-unités de l'anneau supérieur où Csl4, Rrp4 et Rrp40 sont dotés de domaines KH et S1 (Liu *et al.*, 2006).

9.2. Les activités enzymatiques de l'exosome à ARN

Toutes les machineries enzymatiques procaryotiques susmentionnées utilisent un mécanisme d'action phosphorolytique $3' \rightarrow 5'$ afin de dégrader l'ARN (Briani *et al.*, 2016, Evguenieva-Hackenberg *et al.*, 2014). Cette activité dépend de la présence de multiples sites PH actifs (six pour la RNase PH et trois pour la PNPase ou l'exosome archéen) tapissant l'intérieur du canal central généré par la forme toroïde de ces complexes. Chez les eucaryotes, les sites actifs putatifs des sous-unités à domaine *RNase-PH like* de l'exosome à ARN ont toutefois subi des mutations les rendant inactifs (Dziembowski *et al.*, 2007). L'exosome à ARN eucaryotique n'est toutefois pas dépourvu d'activité enzymatique pour autant, ayant évolué de manière à aller chercher des activités enzymatiques via l'établissement d'interactions avec Dis3/Rrp44 et/ou Rrp6 (Exo¹⁰/Exo¹¹), deux exoribonucléases $3' \rightarrow 5'$ (Figure 10).

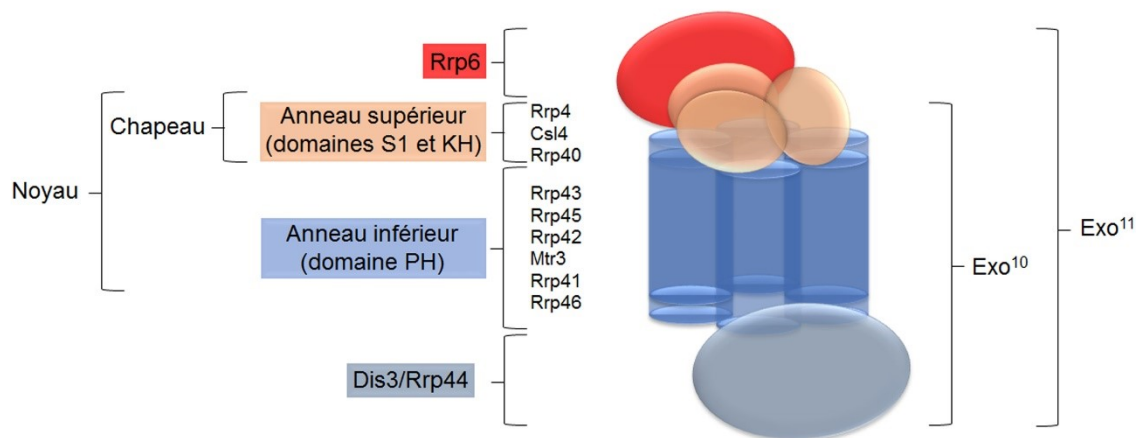


Figure 10. Schéma de l'architecture de l'exosome à ARN.

L'exosome à ARN chez la levure est constitué de 10 sous-unités protéiques (Exo¹⁰). Neuf de ces sous-unités s'assemblent de manière à former une superposition de structures toroïdales qui définissent le noyau de l'exosome à ARN. L'anneau inférieur est composé de six sous-

unités: Rrp43, Rrp45, Rrp42, Mtr3, Rrp41 et Rrp46 chacune possédant un motif structural (domaine PH) retrouvé chez la RNase PH d'*E. coli*. L'anneau supérieure, surnommé le chapeau de l'exosome à ARN, comprend les sous-unités Rrp4, Csl4 et Rrp40 lesquelles sont dotées de domaines de liaison à l'ARN S1 et/ou KH. Le noyau de l'exosome à ARN est catalytiquement inactif. Les activités enzymatiques associées à l'exosome à ARN sont ainsi pourvues par l'entremise de Rrp6, une exoribonucléase 3'→5', ainsi que par Dis3/Rrp44 laquelle possède à la fois une activité exoribonucléase 3'→5' et une activité endonucléase. Rrp6 et Dis3 se positionnent respectivement de part et d'autre du noyau de l'exosome à ARN avec Rrp6 située à son extrémité supérieure. La présence de Rrp6 (Exo¹¹) définit la version nucléaire de l'exosome à ARN chez la levure. Les sous-unités ne sont pas dessinées à l'échelle.

9.2.1. Dis3

Dis3 fait partie de la superfamille d'enzyme RNR qui comprend également la RNase II et la RNase R d'*E. coli* (Zuo et Deutscher, 2001). Structurellement, Dis3 est constituée d'un motif CR3 (*cysteine-Rich with three cysteines motif*), d'un domaine *Pilus-forming N-terminal* (PIN), de deux domaines *cold-shock* (CSD), d'un domaine ribonucléase (RNB) ainsi qu'un domaine S1 (Robinson *et al.*, 2015). Fonctionnellement, les domaines CSDs et S1 sont impliqués dans la reconnaissance non-spécifique à l'ARN (Lorentzen *et al.*, 2008), alors que les domaines PIN en RNB sont pourvus d'activités enzymatiques. Le domaine PIN favorise la dégradation d'ARN par un mécanisme endoribonucléolytique (Lebreton *et al.*, 2008, Schaeffer *et al.*, 2009, Schneider *et al.*, 2009). L'éventail des substrats physiologiques sujets à l'activité endoRNase de Dis3 est néanmoins largement méconnu (Schneider *et al.*, 2012, Tomecki *et al.*, 2014). Outre cette fonction, le domaine PIN, de concert avec le motif CR3, participe à l'association de Dis3 au noyau de l'exosome à ARN. Par une interaction directe avec Rrp41/Rrp45, le domaine PIN positionne ainsi Dis3 à la base de ce dernier (Bonneau *et al.*, 2009, Schaeffer *et al.*, 2012, Schneider *et al.*, 2009). Le domaine RNB est, pour sa part, doté d'une activité 3'→5' exoribonucléase hydrolytique processive (Dziembowski *et al.*, 2007, Mitchell *et al.*, 1997).

9.2.2. Rrp6

Associée à l'exosome à ARN se trouve une deuxième exoribonucléase, Rrp6. Celle-ci est la seule sous-unité du complexe Exo¹¹ non-essentielle à la viabilité cellulaire. Néanmoins, son absence entraîne l'accumulation nucléaire de nombreux types d'ARNs (Schneider *et al.*, 2012). Rrp6 appartient à la superfamille de nucléase à motif DEDD, une famille d'enzyme à activité 3'→5' exoribonucléase hydrolytique distributive (Burkard et Butler, 2000, Zuo et Deutscher, 2001). Outre son domaine RNB lié à son activité exoribonucléase, Rrp6 possède trois domaines additionnels: un domaine HRDC (*Helicase and RNase D Carboxy Terminal*) impliqué dans la formation du module catalytique (Midtgaard *et al.*, 2006), un domaine PMC2NT (*polycystin 2 N-terminal*) nécessaire à l'établissement d'interaction protéique (Stead *et al.*, 2007) ainsi qu'un domaine carboxy-terminale (CTD) lequel établit des interactions avec les sous-unités Csl4, Rrp4, Rrp40, Mtr3 et Rrp43, positionnant ainsi Rrp6 au sommet de l'exosome à ARN (Callahan et Butler, 2008, Makino *et al.*, 2013, Wasmuth *et al.*, 2014).

9.3. Isoformes de l'exosome à ARN chez la levure

L'association conjointe de Dis3 et Rrp6 au noyau de l'exosome à ARN n'est pas invariable et est modulée dépendamment du compartiment cellulaire dans lequel l'exosome à ARN se trouve. Chez les levures, l'exosome à ARN se présente ainsi sous deux isoformes, l'un nucléaire et l'autre cytoplasmique. Pour chacun de ces isoformes, on trouve un assemblage invariable composé du noyau de l'exosome à ARN et de la sous-unité catalytique Dis3 (Exo¹⁰). Les isoformes cytoplasmiques et nucléaires se distinguent l'un de l'autre par la présence de Rrp6, celle-ci étant spécifiquement restreinte au compartiment nucléaire (Allmang *et al.*, 1999).

9.4. Les voies de dégradation/maturation 3'→5' de l'exosome à ARN nucléaire

Le positionnement opposé des exoribonucléases Rrp6 et Dis3 par rapport au noyau de l'exosome à ARN implique des voies d'accès fondamentalement distinctes pour les substrats

d'ARNs. Une de ces voies implique le canal central de l'exosome à ARN. Celui-ci joue un rôle primordial dans la dégradation/maturation de nombreux substrats en dirigeant des ARNs simple brin (ARNsb) vers Dis3. L'importance du canal central est notamment soulignée par l'inhibition marquée des fonctions de l'exosome à ARN *in vitro* et *in vivo* lors de l'introduction de mutations à l'intérieur du canal central générant de l'encombrement stérique ou une diminution de l'affinité de liaison à l'ARN suite à un inversement de charge (Drazkowska *et al.*, 2013, Wasmuth et Lima, 2012). Une telle voie d'accès est également supportée par de nombreuses études structurales (Bonneau *et al.*, 2009, Makino *et al.*, 2013, Wasmuth *et al.*, 2014). Afin d'y avoir recours, l'ARNsb doit minimalement être d'une longueur de 31-33 nucléotides, dont 20 sont nécessaires afin de traverser le canal central, le reste se liant à Dis3 (Bonneau *et al.*, 2009) (Figure 11 – Route #1). En l'absence d'une telle longueur, l'exosome à ARN subit un changement de conformation important caractérisé par le redéploiement spatial du domaine RNB de Dis3 par rapport au noyau de l'exosome à ARN (Liu *et al.*, 2014, Makino *et al.*, 2013). Le changement de structure a pour conséquence d'obstruer l'endroit où l'ARNsb émerge du canal central. Sous une telle conformation, un ARNsb peut néanmoins atteindre le domaine RNB de Dis3 via l'utilisation d'une voie d'accès alternative (voie directe) (Wang *et al.*, 2007) (Figure 11 – Route #2) Celle-ci est notamment utilisée lors la maturation 3' de précurseurs d'ARNt (Liu *et al.*, 2014, Schneider *et al.*, 2012). Les déterminants conduisant de potentiels substrats vers cette voie d'accès restent cependant à être déterminés. Cela dit, les ARNsb empruntant cette voie alternative sont d'une longueur inférieure à ceux empruntant le canal central et sont hautement structurés (Bonneau *et al.*, 2009, Liu *et al.*, 2014, Makino *et al.*, 2013).

Tout comme la voie d'accès alternative de Dis3, l'accès au domaine enzymatique de Rrp6 peut s'effectuer indépendamment du canal central de l'exosome à ARN (Makino *et al.*, 2015) (Figure 11 – Route #3). Or, une deuxième voie d'accès existe également afin d'accéder au domaine catalytique de Rrp6. Dans celle-ci, l'ARNsb entre par la côté de l'exosome à ARN par le biais d'ouvertures créées à l'interface des sous-unités du chapeau et du noyau de l'exosome à ARN. Se faisant, l'ARNsb rejoint le canal central avant de bifurquer et traverser les sous-unités du chapeau de l'exosome à ARN afin d'ultimement rejoindre le site actif de Rrp6 (Wasmuth *et al.*, 2014, Wasmuth et Lima, 2012) (Figure 11 – Route #4). Cette dernière

voie d'accès est en accord avec le fait qu'un exosome à ARN dépourvu de l'activité 3'→5' exoribonucléase de Dis3 inhibe fortement l'activité enzymatique de Rrp6 (Wasmuth et Lima, 2012). Dans ce contexte, l'ARNsb séquestré par la version inactive de Dis3 bloque le canal central de l'exosome à ARN rendant dès lors impossible l'accès à Rrp6.

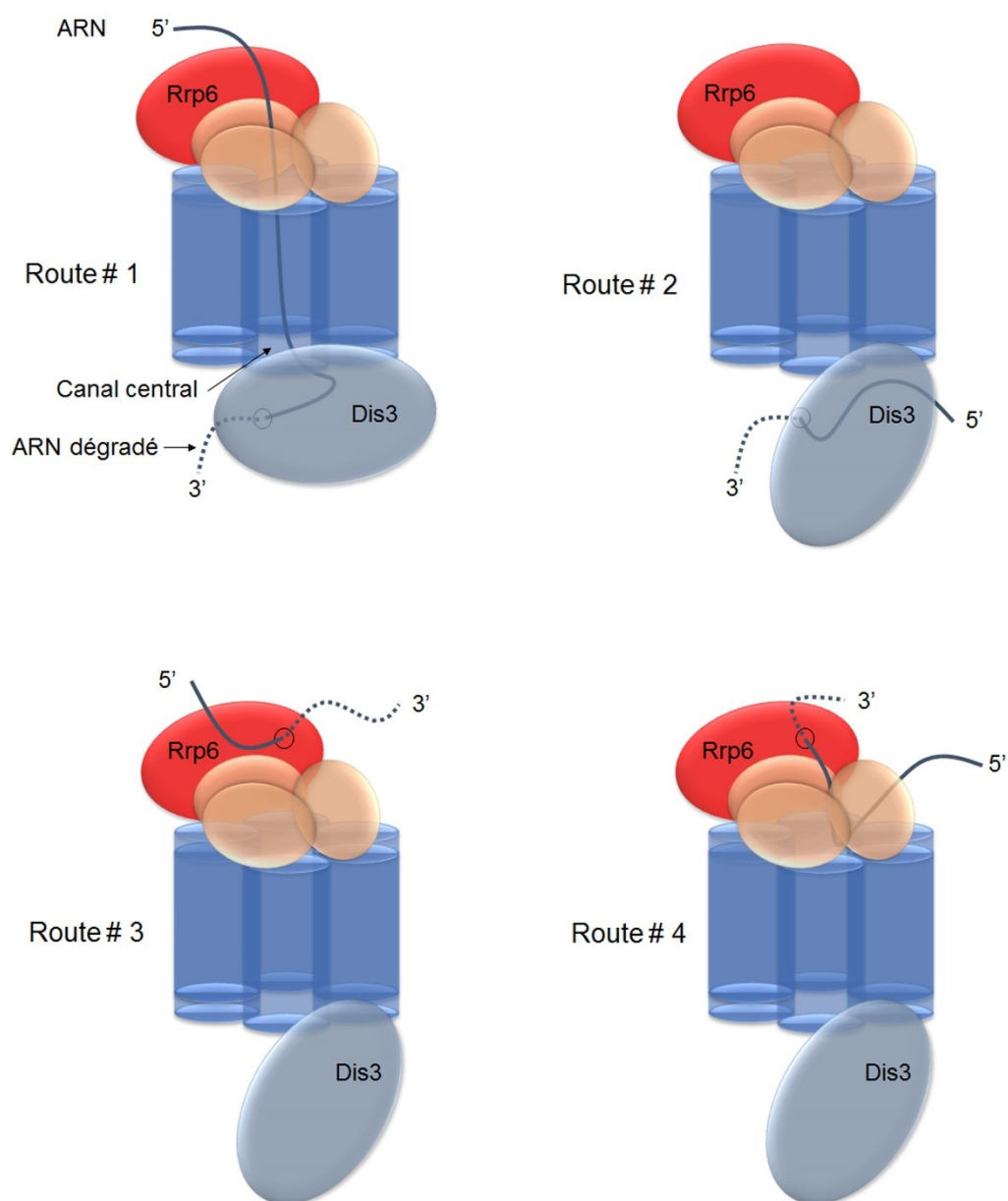


Figure 11. Schéma des voies de dégradation/maturation 3'→5' de l'exosome à ARN nucléaire.

Quatre options sont disponibles à un substrat d'ARN afin d'accéder aux exoribonucléases 3'→5' de l'exosome à ARN nucléaire. Dans la route #1, l'ARN passe à l'intérieur d'un canal central créé par la structure toroïdale de l'exosome à ARN. L'ARN accède au canal central par les protéines du chapeau de l'exosome à ARN puis parcourt la totalité du canal central afin de rejoindre le site actif de Dis3 où il est sujet à une dégradation (ligne pointillée). L'accès au domaine RNB de Dis3 peut également s'effectuer par une voie alternative, surnommée la voie d'accès directe (route #2). Les ARNs empruntant cette voie sont d'une taille inférieure à ceux utilisant la route #1. En fait, la longueur d'un substrat d'ARN est associée à la régulation du positionnement de Dis3 à la base du noyau de l'exosome à ARN. L'absence d'ARN à l'intérieur du canal central de l'exosome à ARN entraîne l'exposition au solvant du domaine RNB de Dis3 générant ainsi une voie d'accès directe à la base de l'exosome à ARN. Deux voies d'accès existent également afin de d'atteindre le site actif de Rrp6. Dans la route #3, le substrat d'ARN accède directement à Rrp6 sans nécessité la participation de l'Exo¹⁰. Alternativement, dans la route #4, l'ARN accède au site catalytique de Rrp6 par un point d'entrée situé à l'interface entre l'anneau inférieur et supérieur du noyau de l'exosome à ARN. Dans cette route, l'ARN emprunte un parcours atypique, cheminant vers l'extrémité supérieure de Exo¹¹ au contraire de la voie #1. La route #4 est la seule voie d'accès non validée par des expériences en contexte biologique. L'exosome à ARN n'est pas dessiné à l'échelle.

10. Les cofacteurs de l'exosome à ARN

Afin de contrôler et/ou augmenter la spécificité et l'activité des événements de maturation et/ou de dégradation, l'exosome à ARN interagit avec un ensemble de cofacteurs tant nucléaires que cytoplasmiques. Ces cofacteurs peuvent servir de protéines adaptatrices entre l'exosome à ARN et des séquences/structures spécifiques au niveau de l'ARN ou encore donner accès au substrat par le biais d'activités enzymatiques (Butler et Mitchell, 2011).

Chez *S. cerevisiae*, un important cofacteur nucléaire de l'exosome à ARN est le complexe TRAMP (*Trf-Air-Mtr4 Polyadenylation*) (Houseley et Tollervy, 2006, LaCava *et al.*, 2005, Vanacova *et al.*, 2005, Wyers *et al.*, 2005). Ce complexe est formé de trois protéines : une protéine pouvant lier l'ARN (Air1/Air2), une poly(A) polymérase non-canonique (Trf4/Trf5) ainsi qu'une hélicase 3'→5' ARN-dépendante de la famille des hélicases à boîte DExH (Mtr4). Le complexe TRAMP facilite l'activité de l'exosome à ARN via l'ajout d'une courte série d'adénosines (entre 3 et 5 nucléotides) à l'extrémité 3' des ARNs ciblés (Jia *et al.*, 2011, Wlotzka *et al.*, 2011). L'ajout de cette courte queue poly(A) permet le recrutement de Mtr4 qui, par son activité hélicase, vient défaire (linéariser/décomplexer) les structures secondaires

au sein du substrat d'ARN favorisant ainsi sa prise en charge par l'exosome à ARN (Bernstein *et al.*, 2010, Bernstein *et al.*, 2008, Jia *et al.*, 2012).

Le complexe TRAMP possède des équivalences chez plusieurs organismes. Chez *S. pombe*, celui-ci est composé de Cid14 (Trf4), Air1 et Mtr4 (Keller *et al.*, 2010). Un complexe TRAMP est également présent au niveau du nucléole des cellules mammifères (hTRAMP) et, semblable aux levures, est constitué d'une poly(A) polymérase non-canonique (Trf4-2/Trf4), d'une protéine à doigts de zinc (ZCCHC7/Air2) ainsi que de l'hélicase hMTR4 (SKIV2L2) (Fasken *et al.*, 2011, Lubas *et al.*, 2011). Cette dernière fait également partie du complexe NEXT (*nuclear exosome targeting complex*), un complexe nucléoplasmique qui agit aussi comme activateur de l'exosome à ARN. Cela dit, la manière dont le complexe TRAMP sélectionne les transcrits dédiés à l'exosome à ARN est largement méconnue. La liaison du complexe TRAMP à l'ARN est non-spécifique étant associée à aucun motif d'ARN particulier (Holub *et al.*, 2012, Wlotzka *et al.*, 2011). Néanmoins, une certaine spécificité peut être obtenue par le biais de protéines adaptatrices. Ainsi, chez *S. cerevisiae*, Nop53 et Utp18, deux facteurs nécessaires à la maturation d'ARNs pré-ribosomiaux, interagissent directement avec Mtr4 (Thoms *et al.*, 2015). Le complexe TRAMP interagit également avec le complexe de terminaison transcriptionnelle NNS via une liaison entre Nrd1 et Trf4 (Tudek *et al.*, 2014). Au sein du complexe NEXT, hMTR4 s'associe avec deux protéines pouvant lier l'ARN, ZCCHC8 et RMB7 et est donc dépourvue d'activité de polyadénylation (Lubas *et al.*, 2011). Ainsi, autant l'activité de polymérase non-canoniques que la reconnaissance d'éléments de séquences caractéristiques par divers cofacteurs peut servir à favoriser l'activité enzymatique de l'exosome à ARN (Rammelt *et al.*, 2011) (Hrossova *et al.*, 2015, Kim *et al.*, 2016, Lubas *et al.*, 2015, Thoms *et al.*, 2015, Tudek *et al.*, 2014). Cela dit, qu'importe la manière utilisée afin de diriger un substrat d'ARN vers l'exosome à ARN, il appert qu'une activité hélicase 3'→5' représente une facette importante de la régulation de l'activité de l'exosome à ARN (Hardwick et Luisi, 2013).

11. Les déterminants dictant le choix entre Dis3 et Rrp6

Le choix d'une exoribonucléase au détriment d'une autre est une question non-résolue, mais d'un grand intérêt. Autant Dis3 que Rrp6 peut être influencée par l'activité du complexe TRAMP (Callahan et Butler, 2010, Schneider *et al.*, 2007). D'après des études de cristallographie, le complexe TRAMP est recruté à l'extrémité supérieure de l'exosome à ARN par le biais d'interactions entre Mtr4 et d'une surface formée de l'assemblage de Rrp6 et de son partenaire d'interaction Rrp47 (Makino *et al.*, 2015, Schuch *et al.*, 2014). Le complexe TRAMP est donc orienté de manière à fournir des ARNs à la fois à Dis3, via l'utilisation du canal central, mais également à Rrp6. Un concept émergent suggère que ce choix est influencé par un changement de conformation chez l'exosome à ARN. Ce changement implique un repositionnement de Rrp6/Rrp47/Mtr4 par rapport à l'Exo¹⁰ ce qui favoriserait un ciblage préférentiel des substrats vers Rrp6 (Makino *et al.*, 2015). Ce modèle serait notamment cohérent avec la maturation de l'ARNr 5.8S. Celui-ci est initialement transcrit sous la forme d'un long précurseur (extension 3') pour être mûré subséquemment par Rrp6 et/ou Dis3. Toutefois, lorsque le précurseur atteint 30 nucléotides (5.8S + 30), Rrp6 devient l'enzyme responsable de sa maturation finale (Allmang *et al.*, 1999). À cette longueur, l'exosome à ARN serait dans l'environnement immédiat de la particule 60S du ribosome. Cet encombrement stérique serait à la base du repositionnement de Rrp6 ce qui permettrait à celle-ci d'accéder plus facilement au précurseur 5.8S + 30. Donc, outre la longueur d'un substrat d'ARN, l'intégrité du contenu ribonucléoprotéique d'un ARN pourrait donc servir à la fois au choix de l'enzyme, mais également à déterminer si un transcrit sera dégradé ou mûré.

12. Hypothèses et objectifs des projets de recherche

Dans un premier temps, mes activités de recherche se sont concentrées sur Pab2, l'orthologue présumé chez *S. pombe* de PABPN1. Alors que les PABPs cytoplasmiques influencent la stabilité et la traduction de l'ARNm, PABPN1, d'après des études *in vitro*, stimulerait la réaction de polyadénylation en plus de contrôler la taille ultime des queues poly(A). Cependant, la ou les fonction(s) biologique(s) de PABPN1 est/sont largement

méconnue(s). Chez *S. pombe*, l'absence de Pab2 entraîne la présence de transcrits hyperadénylés. Prenant en considération l'impact de la queue poly(A) sur la biologie d'un ARN et ses fonctionnalités, une hypothèse de recherche a été émise laquelle prédit une modulation de l'expression des transcrits hyperadénylés. Les objectifs de ce premier projet de recherche ont donc été d'identifier les ARNs déployant un changement d'expression en absence de Pab2, d'établir une corrélation entre les changements d'expression génique et le profil d'hyperadénylation de cellules *pab2Δ* et, enfin, de caractériser le mécanisme moléculaire Pab2-dépendant limitant l'hyperadénylation.

L'aboutissement du premier projet de recherche a mené à l'identification d'un rôle pour Pab2 lors de la biogénèse des *snoRNAs*. Chez *S. cerevisiae*, les *snoRNAs* sont générés à partir de longs précurseurs polyadénylés par l'activité 3'→5' exonucléase de l'exosome à ARN dans un processus qui est intimement lié au mécanisme utilisé pour terminer leur transcription. Celui-ci implique la participation du complexe trimérique NNS (Nrd1-Nab3-Sen1). Chacune de ces composantes possède un orthologue putatif chez *S. pombe*. En considération de ce point et du fait que l'exosome à ARN est impliqué dans la formation des *snoRNAs*, l'hypothèse de ce deuxième projet de recherche veut qu'un complexe similaire au complexe NNS de chez *S. cerevisiae* soit présent et fonctionnel lors de la terminaison de la transcription des *snoRNAs* chez *S. pombe*. L'objectif de ce projet de recherche a ainsi été d'établir la contribution du présumé complexe NNS de *S. pombe* lors de la terminaison transcriptionnelle des *snoRNAs*.

Mon troisième projet de recherche s'est pour sa part concentré sur l'exosome à ARN. L'exosome à ARN est un complexe multiprotéique évolutionnairement conservé qui participe à la régulation de l'expression génique en ciblant pour dégradation des transcrits aberrants. L'idée communément acceptée est que cette activité s'effectue de manière post-transcriptionnelle. Toutefois, chez plusieurs organismes, l'exosome à ARN est présent à la chromatine de gènes transcriptionnellement actifs (Andrulis *et al.*, 2002, Basu *et al.*, 2011, Hessle *et al.*, 2009, Hieronymus *et al.*, 2004, Lim *et al.*, 2013). L'hypothèse de ce projet de recherche veut donc que l'exosome à ARN participe à un mécanisme de surveillance co-transcriptionnel des ARNs. Les objectifs de ce projet de recherche ont donc été d'analyser

l'association de l'exosome à ARN à la chromatine ainsi que d'évaluer si la modulation du transcriptome induite par la déplétion de l'exosome à ARN était liée à une altération du profil d'occupation de l'ARNPII au niveau des gènes.

Globalement, les travaux de recherche présentés dans cette thèse s'alignent autour d'un thème centré sur les événements de maturation 3' de l'ARN et de la terminaison de la transcription médiée par l'ARNPII. En contrepartie des phases d'initiation transcriptionnelle et d'élongation, la phase de maturation 3'/terminaison fait figure de parent pauvre sur le plan mécanistiques. Quoique qu'initialement non reliés par un fil conducteur clair, les trois projets présentés au sein de cette thèse ont su, au final, s'imbriquer les uns aux autres apportant ultimement de nouvelles informations sur les bases mécanistiques sous-jacentes aux processus essentiels que sont la maturation 3' des ARNs et la terminaison transcriptionnelle de l'ARNPII à l'extrémité 3' des gènes codants et non-codants.

ARTICLE 1**THE NUCLEAR POLY(A)-BINDING PROTEIN INTERACTS
WITH THE EXOSOME TO PROMOTE SYNTHESIS OF NONCODING
SMALL NUCLEOLAR RNAS**

Jean-François Lemay¹, Annie D'Amours¹, Caroline Lemieux¹, Daniel H. Lackner²,
Valérie G. St-Sauveur¹, Jürg Bähler², and François Bachand¹

Publié dans: *Molecular Cell* (2010) 37: 34-45

¹ RNA Group, Université de Sherbrooke, Department of Biochemistry, Québec, Canada

² Research Department of Genetics, Evolution and Environment and UCL Cancer Institute,
University College London, U.K.

CONTRIBUTIONS

Ce projet est le produit d'une collaboration fructueuse entre les laboratoires de recherche des professeurs Bähler et Bachand.

Les validations, par buvardage de type *Northern*, des résultats des puces à ADN, les essais de croissance, l'analyse des expériences de ChIP par qPCR ainsi que les expériences de RIP furent sous ma responsabilité. J'ai également contribué à la création de quelques souches de levures. En collaboration avec F.B., j'ai conçu les diverses expériences du projet et élaboré le modèle du rôle de Pab2 dans la synthèse des *snoRNAs* chez *S. pombe*. De plus, j'ai mis sur pied et exécuté la quasi-totalité des expériences requises suite à la demande de révision du manuscrit. Somme toute, j'ai donc contribué significativement à l'avancement et à l'aboutissement de ce projet. Sommaire des figures/expériences réalisées par J-F.L. : Fig. 2; Fig. 5; Fig. 6C-D; Fig. 7; Fig. S7; Fig. S8.

Le groupe de recherche du professeur Bähler se spécialise notamment dans l'étude bio-informatique des changements dynamiques des programmes d'expression génique suite à des perturbations génétiques. D.H.L, sous la supervision de J.B., fut donc responsable de l'exécution des expériences d'expression par puce à ADN ainsi que de l'analyse des données découlant de ces expériences. D.H.L. est également responsable pour toute autre analyse bio-informatique effectuée lors de cette étude. Les expériences de microscopie à fluorescence ont été effectuées par A.D et V.G.S. A.D. est également responsable des expériences de co-immunoprécipitation *in vivo* et *in vitro*. C.L. a effectué les expériences de ChIP. F.B. a supervisé le projet et écrit le manuscrit.

RÉSUMÉ

Les protéines liant une succession de ribonucléotides de type adénosine, communément appelé queue poly(A), ont une influence majeure sur l'expression génique chez les eucaryotes. Chez l'humain, la version nucléaire de ces PABP, PABPN1, favorise, à l'extrémité 3' des précurseurs d'ARNm, une formation rapide de la queue poly(A) en plus d'exercer un contrôle sur la taille de celle-ci. Chez *S. pombe*, la délétion du gène *pab2*, l'orthologue putatif de PABPN1, entraîne l'accumulation de transcrits hyperadénylés. L'identité de ces transcrits ainsi que le mécanisme menant à leur hyperadénylation sont toutefois inconnus. Ici, nous identifions comme ARN hyperadénylés les petits ARN nucléolaires (*snoRNAs*), lesquels sont typiquement dépourvus de queue poly(A), et caractérisons le mécanisme impliqué dans cette transition. Notamment, nos données révèlent que Pab2 favorise la formation de *snoRNA* matures non-polyadénylés en promouvant la dégradation de la queue poly(A) de précurseurs hyperadénylés via le recrutement de l'exosome à ARN nucléaire. Nos résultats caractérisent donc une fonction inattendue pour Pab2 dans la synthèse d'ARN non-codants, ce qui contraste avec l'idée communément acceptée que les PABP agissent exclusivement qu'au niveau des transcrits codants pour des protéines, et ouvre de nouvelles perspectives sur le mécanisme de recrutement de l'exosome à ARN à des transcrits polyadénylés.

ABSTRACT

Poly(A)-binding proteins (PABPs) are important to eukaryotic gene expression. In the nucleus, the poly(A)-binding protein PABPN1 is thought to function in polyadenylation of pre-mRNAs. Deletion of fission yeast *pab2*, the homolog of mammalian *PABPN1*, results in transcripts with markedly longer poly(A) tails, but the nature of the hyperadenylated transcripts as well as the mechanism that leads to RNA hyperadenylation remain unclear. Here we report the unanticipated finding that Pab2 functions in the synthesis of noncoding RNAs, contrary to the notion that poly(A)-binding proteins function exclusively on protein-coding mRNAs. Accordingly, the absence of Pab2 leads to the accumulation of polyadenylated small nucleolar RNAs (snoRNAs). Our findings suggest that Pab2 promotes poly(A) tail trimming from pre-snoRNAs by recruiting the nuclear exosome. This work unveils a previously uncharacterized function for the nuclear poly(A)-binding protein in snoRNA synthesis and provide new insights into exosome recruitment to polyadenylated RNAs.

INTRODUCTION

Polyadenylation of RNA is fundamental to posttranscriptional gene regulation. In eukaryotes, the 3' poly(A) tail is generally thought to confer positive roles in the mRNA life cycle, such as nuclear export competence, stability, and translational activity. In contrast to these positive roles, recent studies in humans, plants, and yeast reveal that the poly(A) tail can also target an RNA for degradation via the nuclear exosome (Chekanova *et al.*, 2007, LaCava *et al.*, 2005, Vanacova *et al.*, 2005, Wang *et al.*, 2008, West *et al.*, 2006, Wyers *et al.*, 2005).

The exosome complex is one of the primary systems that operates in RNA processing and degradation in eukaryotes. Although the exosome is evolutionarily conserved, it has been most extensively studied in *Saccharomyces cerevisiae*. In this organism, the exosome is composed of nine catalytically-inactive polypeptides to which is associated the active 3'→5' exonuclease, Rrp44/Dis3 (Houseley *et al.*, 2006, Ibrahim *et al.*, 2008, Lebreton et Seraphin, 2008, Schmid et Jensen, 2008, Vanacova et Stefl, 2007). Because inactivation of the exosome via depletion or mutation of any of these ten subunits perturbs similar steps in RNA processing, they are often referred to as the core exosome. The core exosome is found in the nucleus and the cytoplasm. In the nucleus, the activity of the core exosome is associated with an additional 3'→5' exonuclease, Rrp6 (Allmang *et al.*, 1999). Despite the physical association between Rrp6 and the nuclear exosome, genetic and biochemical evidence indicate that they can perform distinct roles in RNA processing (Allmang *et al.*, 1999, Callahan et Butler, 2008, Grzechnik et Kufel, 2008, Milligan *et al.*, 2005, Mitchell *et al.*, 2003, van Hoof *et al.*, 2000).

Polyadenylation-dependent exosome processing can lead to the complete digestion or the partial 3'-end trimming of a substrate RNA (Houseley *et al.*, 2006, Lebreton et Seraphin, 2008, Schmid et Jensen, 2008, Vanacova et Stefl, 2007). As yet, the molecular details behind the choice between 3'-end maturation and complete degradation by the exosome are not clearly understood. Poly(A)-dependent exosome activation is promoted by a polyadenylation complex called TRAMP (for *Trf4/5-Air1/2-Mtr4* polyadenylation complex).

Notably, the polymerase activity of the TRAMP complex is not mediated by the canonical nuclear poly(A) polymerase (Pap1 in *S. cerevisiae*) that catalyzes mRNA polyadenylation, but via the product of the *TRF4* or *TRF5* genes in *S. cerevisiae* (Houseley et Tollervey, 2006, Vanacova *et al.*, 2005, Wyers *et al.*, 2005) and the *cid14* gene in *S. pombe* (Win *et al.*, 2006). Although TRAMP-mediated polyadenylation is believed to mark most RNAs that are destined for exosome-dependent processing, the accumulation of specific polyadenylated transcripts in *RRP6D* cells reduces after inactivation of Pap1 (Carneiro *et al.*, 2007, Grzechnik et Kufel, 2008, van Hoof *et al.*, 2000), suggesting a role for the canonical poly(A) polymerase in exosome-dependent processing.

The importance of exosome-dependent gene regulation is highlighted by the many different types of nuclear RNAs that are targeted by this pathway, including aberrant pre-mRNAs (Bousquet-Antonelli *et al.*, 2000, Milligan *et al.*, 2005, Saguez *et al.*, 2008, West *et al.*, 2006), unprocessed rRNAs (Allmang *et al.*, 1999, Allmang *et al.*, 2000, Dez *et al.*, 2006), unmodified tRNAs (Kadaba *et al.*, 2006, Vanacova *et al.*, 2005), cryptic unstable transcripts (Arigo *et al.*, 2006, Vasiljeva *et al.*, 2008, Wyers *et al.*, 2005), as well as small nuclear (sn) and small nucleolar (sno) RNAs (Allmang *et al.*, 1999, Grzechnik et Kufel, 2008, van Hoof *et al.*, 2000). snoRNAs are an important class of nonpolyadenylated RNAs that assemble into ribonucleoprotein (RNP) particles to act in pre-rRNA processing and modification (Reichow *et al.*, 2007). snoRNAs can be divided into two functional classes depending on the type of pre-rRNA modification that is promoted: box C/D snoRNAs guide 2'-O-ribose methylation, whereas box H/ACA snoRNAs guide pseudouridylation. While the majority of yeast snoRNAs are independently transcribed as larger precursors by RNA polymerase II, a limited number are encoded from intronic sequences of pre-mRNAs (Kiss *et al.*, 2006). A polyadenylation-independent maturation pathway that involves 3'→5' trimming of 3'-extended precursors by the exosome has been proposed for the synthesis of independently-transcribed yeast snoRNAs (Carroll *et al.*, 2004, Fatica *et al.*, 2000, Kim *et al.*, 2006, Steinmetz *et al.*, 2001, Vasiljeva et Buratowski, 2006). However, a recent study suggests that a polyadenylation step is important for pre-snoRNA 3'-end processing via the nuclear exosome (Grzechnik et Kufel, 2008). Yet, the mechanism by which 3' poly(A) tails recruit the nuclear exosome to specific transcripts remains poorly understood.

The evolutionarily conserved nuclear poly(A)-binding protein (PABP2/PABPN1) is the nuclear counterpart of the well-studied cytosolic PABP (PABPC1 in humans and Pab1 in yeast). PABPN1 is characterized by a putative coiled-coil region, a single RNA recognition motif, and a carboxy-terminal glycine/arginine-rich domain (Kuhn et Wahle, 2004, Mangus *et al.*, 2003). Experiments using *in vitro* assays suggest that PABPN1 plays essential roles in mRNA polyadenylation: (i) PABPN1 stimulates processive poly(A) synthesis by direct and simultaneous interactions with the poly(A) polymerase and the growing poly(A) tail (Kerwitz *et al.*, 2003) and (ii) PABPN1 promotes, via a poorly-defined mechanism, the transition from processive to distributive synthesis after addition of approximately 200-300 adenine residues (Wahle, 1995). We have recently reported the identification of a PABPN1 homolog in the fission yeast *Schizosaccharomyces pombe* (Perreault *et al.*, 2007). Notably, deletion of *S. pombe pab2* results in the expression of RNAs with hyperadenylated tails (Perreault *et al.*, 2007); yet, the mechanism that leads to hyperadenylation remains unknown.

Here we used DNA microarrays to identify hyperadenylated transcripts detected in *pab2D* cells. Using this genome-wide approach, we found that the absence of Pab2 leads to the accumulation of polyadenylated snoRNAs and we describe the molecular mechanism of their accumulation. We find that Pab2 is recruited to the 3'-end of snoRNA-encoding genes, is bound to polyadenylated snoRNAs, and is physically associated with the nuclear exosome, suggesting that Pab2 recruits the exosome to promote 3'-end processing of polyadenylated pre-snoRNAs. Our results have therefore unveiled an example of a poly(A)-binding protein involved in the synthesis of noncoding RNAs.

RESULTS

Small nucleolar RNAs are upregulated in *pab2*Δ cells

We previously reported that *pab2*Δ cells display hyperadenylated RNAs (Perreault *et al.*, 2007). In principle, poly(A) tail extension could influence RNA metabolism and modulate gene expression. Accordingly, a genome-wide strategy was established to investigate for gene expression changes in *pab2*Δ cells to distinguish between a general or a specific effect of Pab2 on RNA polyadenylation. Reverse transcription steps prior to microarray analysis were randomly primed to prevent any bias due to poly(A) tail length differences between wild-type and *pab2*Δ cells. A global comparison between wild-type and *pab2*Δ cells is presented in Fig. 1A. Notably, the expression levels of most genes were unaffected by the deletion of *pab2*. A statistical analysis was used (Tusher *et al.*, 2001) to identify RNAs that significantly changed their expression levels between wild-type and *pab2*Δ cells. Using a one-class comparison, we identified 113 and 85 genes that demonstrated increased and decreased RNA levels, respectively, in *pab2*Δ cells (Fig. 1A and Table S1). Among these 198 misregulated RNAs are protein-coding genes that could be directly or indirectly regulated by Pab2. Interestingly, several noncoding small nucleolar RNAs (snoRNAs) also showed increased RNA levels in *pab2*Δ cells (Fig. 1B). Because the role of poly(A)-binding proteins in the control of noncoding RNA expression had not previously been described, we decided to address the mechanism by which snoRNAs were upregulated in the absence of Pab2.

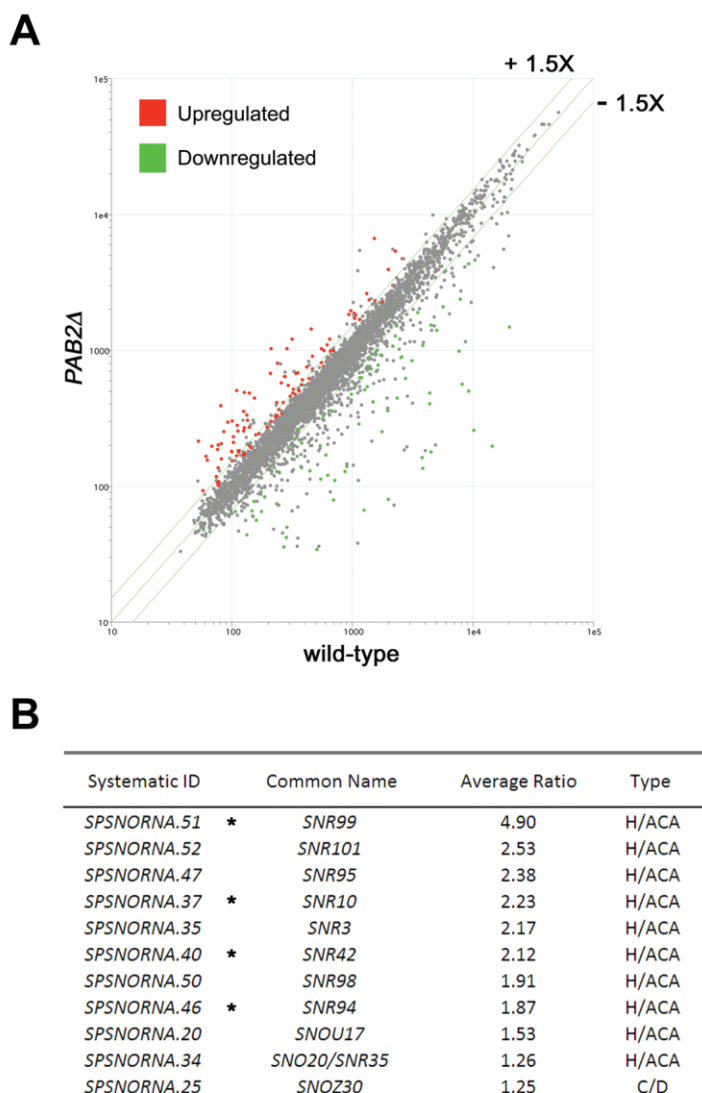


Figure 1. Gene-specific changes in expression level in the *pab2*Δ strain.

(A) Scatter plot of the RNA signals from wild-type (*x*-axis) and *pab2*Δ (*y*-axis) cells. Red and green dots represent genes that demonstrated statistically significant upregulation and downregulation, respectively, in *pab2*Δ cells as determined by the Significance Analysis of Microarray (SAM). (B) Upregulated snoRNAs as determined by expression profiling of *pab2*Δ cells. Asterisks indicate those snoRNAs identified by SAM (see Methods section for details).

***pab2*Δ cells accumulate 3'-extended polyadenylated snoRNAs**

To independently validate the results obtained by DNA microarrays, we compared snoRNA levels between wild-type and *pab2*Δ strains by Northern blotting. Surprisingly, larger heterogeneous population of transcripts were specifically detected in the *pab2*Δ strain along with decreased levels of mature snoRNAs (data not shown). Given that we previously reported an hyperadenylation phenotype in *pab2*Δ cells (Perreault *et al.*, 2007), we suspected that this heterogeneous RNA population could represent polyadenylated snoRNAs. To test this possibility, the 3'-end of several snoRNAs was examined by treating total RNA prepared from wild-type and *pab2*Δ cells with RNase H. RNase H treatment in the presence of a DNA oligonucleotide complementary to a region roughly 100-nt upstream from the mature 3'-end will release a 3' fragment that can be detected by Northern blot. As can be seen in Fig. 2A, RNase H reactions using an oligo specific to H/ACA type snR99 yielded a 100-nt fragment corresponding to the 3'-end of mature snR99. Consistent with our Northern blot assays, RNase H analyses from the *pab2*Δ strain showed heterogeneous 3'-extended snR99 products as well as decreased levels of mature snR99 (Fig. 2A, lane 4 and Fig. 2I). Importantly, addition of oligo d(T) to the RNase H reaction caused the heterogeneous population of 3'-extended snR99 transcripts to migrate as a discrete product (Fig. 2A, lane 3). The collapse of heterogeneous 3'-extended snR99 transcripts into a discrete product after the addition of oligo d(T) suggests that 3'-extended snR99 are polyadenylated. This possibility was confirmed by sequencing 3'-extended snR99 transcripts after conversion to cDNA. DNA sequencing indicated the presence of non-encoded oligo(A) stretches 78-nt downstream from the mature 3'-end of snR99 (data not shown). A distance of 78-nt between the 3'-end of mature snR99 and the start of poly(A) tails is consistent with the detection of a ~178-nt product after RNase H cleavage using both d(T)₁₈ and snR99-specific oligos (Fig. 2A, lane 3).

We next performed similar experiments to test whether polyadenylated products as seen for snR99 were observed for other snoRNAs in *pab2*Δ cells. Notably, all of the independently-transcribed snoRNAs that were tested in *pab2*Δ cells accumulated oligo d(T)-sensitive 3'-extended heterogeneous products (Figs. 2A-2E). The accumulation of

polyadenylated snoRNAs in the *pab2Δ* strain was also associated with reduced levels of several mature snoRNAs (Figs. 2A-2E and 2I). This analysis included two additional H/ACA box snoRNAs (snR3 and snR42; Fig. 2, B & C) as well as two C/D box snoRNAs that were not initially detected in the microarray experiments (snoR56 and snoR68; Fig. 2, D & E). In contrast to independently-transcribed snoRNAs, 3'-end analysis of an intron-embedded C/D box snoRNA, snoR54, did not result in the accumulation of 3'-extended forms in *pab2Δ* cells (Fig. 2F, lanes 3-4). As a control, a 3'-extended form of intron-embedded snoR54 accumulated in cells deleted for *rrp6* (Fig. 2F, lanes 5-6), consistent with previous studies in budding yeast (Allmang *et al.*, 1999, van Hoof *et al.*, 2000). We also analyzed the 3'-end of several mRNAs (Figs. 2G-2H and data not shown). As can be seen for the *adh1* and *pyk1* mRNAs, no differences in 3'-end decision and poly(A) tail length were detected between wild-type and *pab2Δ* cells. These experiments indicate that the absence of Pab2 leads to the accumulation of 3'-extended polyadenylated snoRNAs.

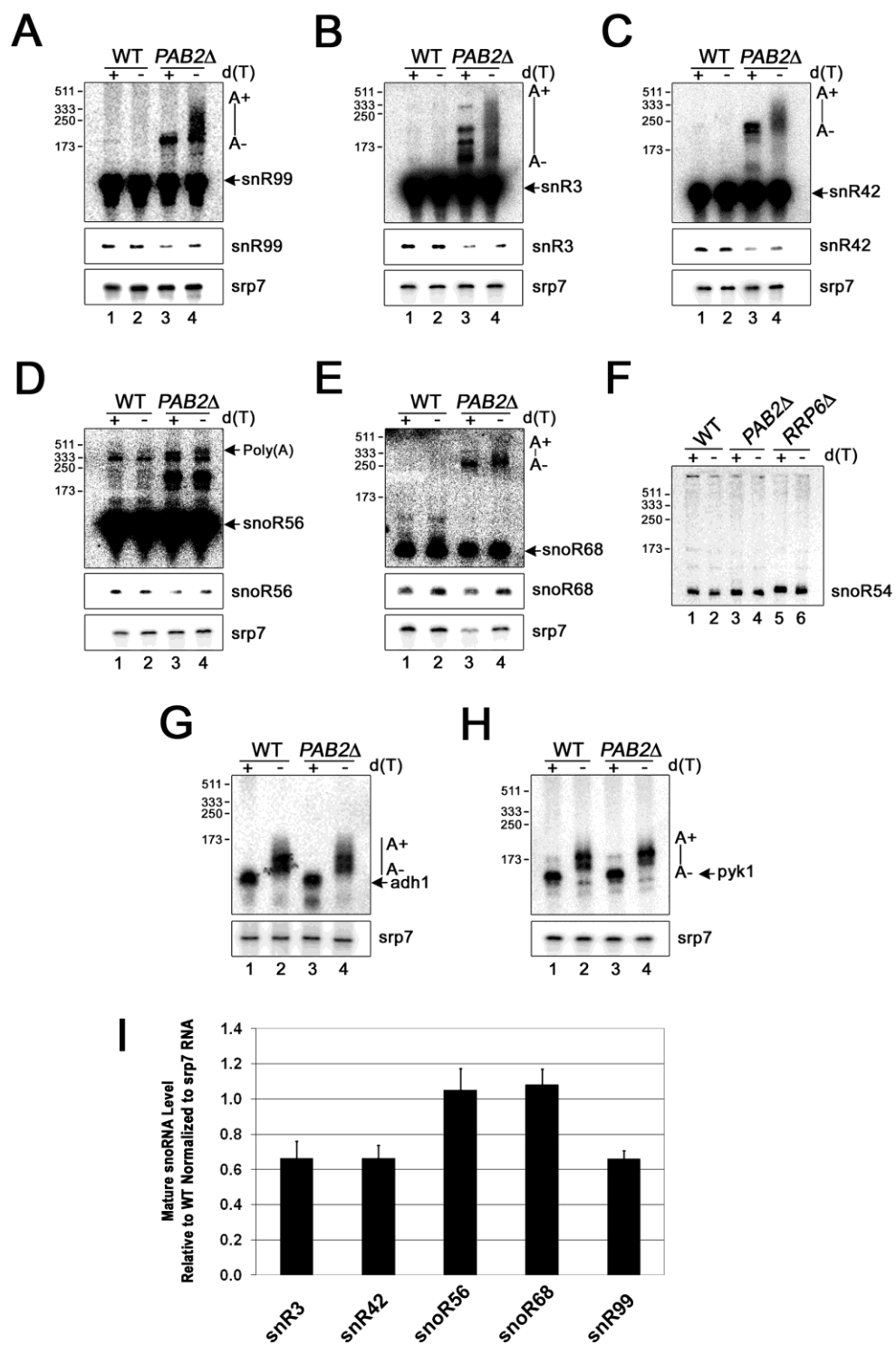
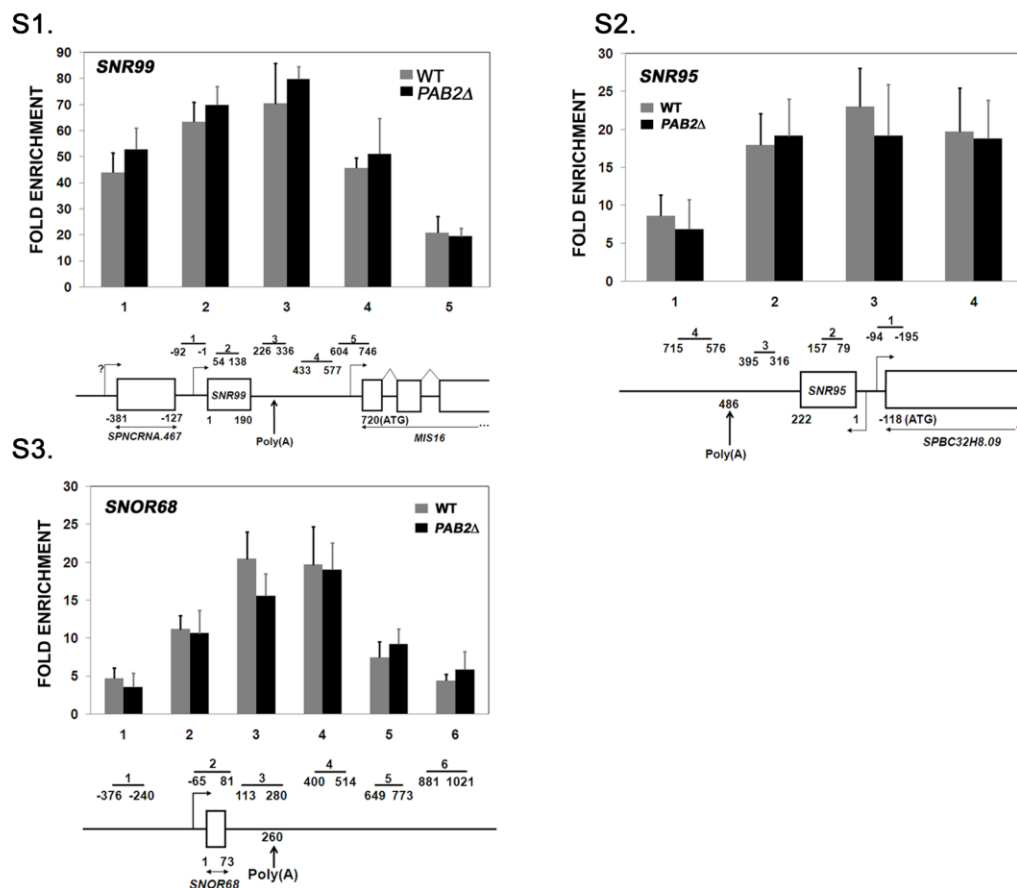


Figure 2. Accumulation of 3'-extended polyadenylated snoRNAs in *pab2Δ* cells.

Total RNA prepared from wild-type and *pab2Δ* cells was treated with RNase H in the presence of DNA oligonucleotides complementary to H/ACA class snoRNAs snR99 (A), snR3 (B), and snR42 (C); C/D class snoRNAs snoR56 (D) and snoR68 (E); intron-derived snoR54 (F); and *adh1* (G) and *pyk1* (H) mRNAs. RNase H reactions were performed in the presence (+) or absence (-) of oligo(dT). *srp7* RNA was used as a loading control. Size markers (in nucleotides) are indicated on the left. (I) Quantification of alterations in mature snoRNA levels, with the wild-type ratio set to 1. Values represent the means of at least three independent experiments and bars correspond to standard deviations.

The accumulation of 3'-extended snoRNAs in *pab2Δ* cells could be the result of increased transcriptional readthrough as a consequence of defects in transcription termination. To address this possibility, we performed chromatin immunoprecipitation (ChIP) assays to examine the density of RNA Pol II along snoRNA genes. ChIP assays have previously been used in yeast to show 3'-extended Pol II crosslinking in mutants defective in transcription termination (Kim *et al.*, 2006, Steinmetz *et al.*, 2006, Thiebaut *et al.*, 2006). Pol II densities as determined by ChIP were similar in regions 3' of the H/ACA class *SNR99* gene between wild-type and *pab2Δ* strains (Fig. S1). Comparable results were observed for two other snoRNA genes with polyadenylation sites 187-nt and 264-nt downstream from the mature snoRNA 3'-end (Fig. S2-S3). The ChIP assays suggest that the accumulation of 3'-extended snoRNAs in *pab2Δ* cells results from defects in a posttranscriptional mechanism, rather than the increased production of readthrough transcripts.



Figures S1-S3. Density of RNA Pol II along snoRNA genes in WT and *pab2Δ* cells.

Chromatin immunoprecipitation (ChIP) assays were performed on extracts prepared from wild-type and *pab2Δ* cells using a monoclonal antibody specific to RNA PolIII. The co-precipitating DNA was quantified by real-time PCR using gene-specific primer pairs located along the *snR99* (S1), *snR95* (S2), and *snoR68* (S3) genes. All of the ChIP data are presented as the fold enrichment of RNA Pol II relative to a control intergenic region. Values represent the means of at least three independent experiments and bars correspond to standard deviations.

3'-extended snoRNAs accumulate in discrete foci in *pab2Δ* cells

To determine the subcellular distribution of 3'-extended snoRNAs that accumulate in *pab2Δ* cells, a 54-nt probe was designed to specifically detect the 3'-extended form of H/ACA class snR99 by fluorescent in situ hybridization (FISH). Analysis of wild-type cells using this 54-nt probe produced background signal (Fig. 3A, panel a), similar to control cells without the fluorescent probe (data not shown). This is consistent with the low level of 3'-

extended snR99 detected in wild-type cells using RNase H assays (Fig. 2A). In contrast, discrete foci were detected in *pab2Δ* cells using the fluorescent probe specific to 3'-extended snR99 (Fig. 3A, panel b). Notably, these foci precisely coincided with poly(A)⁺ RNA-rich foci (Fig. 3A, panels b-d) that are specifically detected in the absence of Pab2 (see below). This is consistent with the accumulation of 3'-extended polyadenylated snR99 in a discrete compartment of *pab2Δ* cells.

To further characterize the cellular localization of foci containing 3'-extended snR99 and poly(A)-rich RNA, oligo d(T) staining was combined with the DNA-intercalating agent DAPI that stains the nucleoplasm as well as with a GFP-tagged version of fibrillarin that localizes to the nucleolus. As can be seen in Fig. 3B, poly(A)-rich foci were detected in *pab2Δ* cells (panel e), but not in wild-type cells (panel a). Comparison of the different staining methods showed that the poly(A) bodies were concentrated in a region that was distinct from the DAPI staining, but that colocalized with GFP signal (Fig. 3B, panels f-h). Such a specific co-localization between the poly(A) foci and Fib1-GFP was detected in the majority of *pab2Δ* cells. The results of the FISH experiments indicate that the nucleolus of *pab2Δ* cells accumulates polyadenylated RNAs, including 3'-extended forms of a snoRNA.

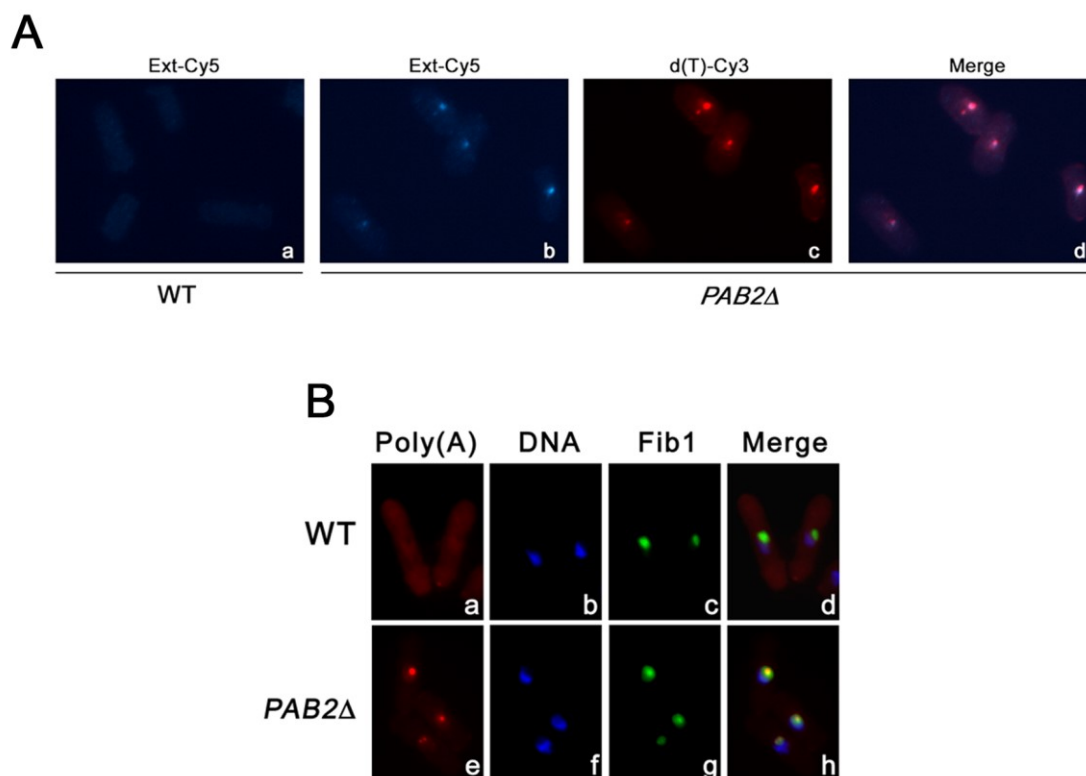


Figure 3. 3'-extended snoRNAs accumulate in discrete foci in *pab2*Δ cells.

(A) Wild-type (panel a) and *pab2*Δ cells (panels b-d) were analyzed by FISH after double labelling with a Cy5-labeled probe for 3'-extended snR99 (panels a-b) and a Cy3-labeled oligo(dT)₅₀ (panel c). (B) Wild-type (panels a-d) and *pab2*Δ (panels e-h) cells expressing GFP-tagged Fib1 were visualized for polyadenylated RNA using Cy3-labeled oligo(dT)₅₀ (panels a and e), for DNA using DAPI (panels b and f), and for Fib1 using GFP (panels c and g).

Functional and physical associations between Pab2 and the nuclear exosome

Given the similar levels in Pol II density downstream of snoRNA genes (Fig. S1-S3), the accumulation of 3'-extended polyadenylated snoRNAs in the *pab2*Δ strain suggested that Pab2 functions via a posttranscriptional mechanism. To get insights into the pathway associated with Pab2-dependent snoRNA synthesis, we compared the expression profile of *pab2*Δ cells with the RNA profiles from all available fission yeast mutants to verify for significant overlap in regulated genes. Significantly, of those 113 Pab2-upregulated genes, 37 (33%) showed increased expression in cells deleted for the 3'→5' exonuclease *rrp6* (Wilhelm *et al.*, 2008) ($P \sim 8 \times 10^{-11}$), and similarly, 37 genes (33%) showed increased

expression in the *dis3-54* strain (Wang *et al.*, 2008) ($P \sim 5 \times 10^{-15}$) (Figs. 4A and 4B). The *dis3-54* strain contains an amino acid substitution in the exonuclease domain of the exosome component Dis3/Rrp44 that impairs its catalytic activity (Murakami *et al.*, 2007). Overall, 24 genes showed overlap between the lists of upregulated genes from *pab2Δ*, *rrp6Δ*, and *dis3-54* (Fig. S4). These results underscore the important functional relationship among Pab2, Rrp6, and Dis3.



Figure 4. Functional and physical associations between Pab2 and the nuclear exosome. Venn diagrams present the numbers of overlapping genes from the lists of genes that were upregulated ≥ 1.5 fold in the *pab2Δ* and *rrp6Δ* strains (A) and in the *pab2Δ* and *dis3-54* strains (B). The P -values indicate the probabilities that the observed overlaps occurred by chance.

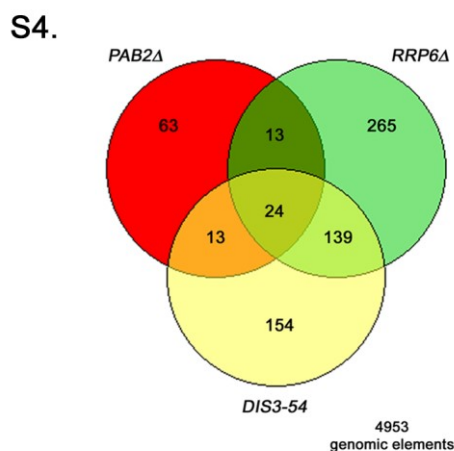


Figure S4. Overlap of upregulated genes from *pab2Δ*, *rrp6Δ*, and *dis3-54* cells. Venn diagrams present the numbers of overlapping genes from the lists of genes that were significantly upregulated in the *pab2Δ*, *rrp6Δ*, and *dis3-54* strains.

The significant overlap between the expression profiles of *pab2Δ*, *rrp6Δ*, and the *dis3-54* mutants suggested that Pab2 might physically interact with components of the exosome complex. To test this possibility, we affinity purified TAP-tagged versions of Rrp6 and Dis3 from extracts of cells that also expressed a HA-tagged version of Pab2. As can be seen in Fig. 4C, Pab2 was co-immunoprecipitated with Dis3-TAP (lane 5) as well as with Rrp6-TAP (lane 6), but was not detected in a control purification (lane 4). These protein associations were not sensitive to RNases (data not shown). These results indicate that Rrp6 and Dis3 can be found in a complex with Pab2.

The results of the affinity purification experiments indicated that the enrichment of Pab2 in the Rrp6 purification was 16-fold greater than with Dis3 (Fig. 4C). We therefore examined whether Pab2 makes direct contact with Rrp6. For this, the two proteins were expressed in *E. coli* and the ability of purified Rrp6 to interact with GST-Pab2 was analyzed *in vitro*. Figure 4D shows that Rrp6 effectively bound to GST-Pab2 (lane 5), but not GST alone (lane 6). Similar results were obtained using soluble Pab2 and Rrp6 immobilized on beads (Fig. S5). These experiments indicate that Pab2 directly interacts with Rrp6.

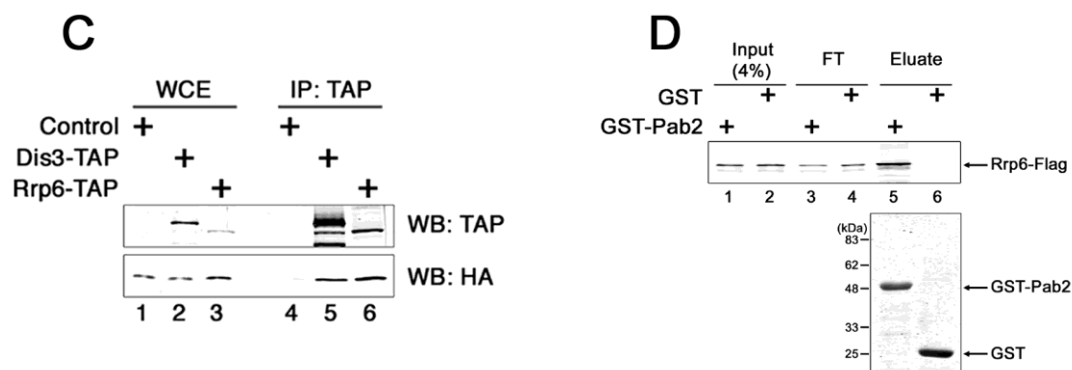


Figure 4. Functional and physical associations between Pab2 and the nuclear exosome. (C) Immunoblot analysis of whole cell extracts (WCE; lanes 1-3) and IgG-sepharose precipitates (IP; lanes 4-6) prepared from control, Dis3-TAP, and Rrp6-TAP cells expressing a HA-tagged version of Pab2. (D) Direct interaction between Pab2 and Rrp6 *in vitro*. Equal amounts of purified Rrp6 were incubated with GST and GST-Pab2. Western blot (anti-Flag; upper panel) and corresponding Coomassie blue-stained loading control (lower panel) are shown. *FT*, Flow-through.

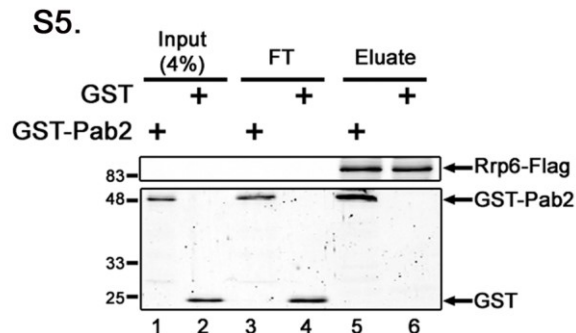


Figure S5. Rrp6 directly interacts with Pab2.

Equal amounts of purified GST (lanes 1, 3, and 5) and GST-Pab2 (lanes 2, 4, and 6) were incubated with purified Rrp6 immobilized on beads. Following a 1-hr incubation, the flow-through (FT; lanes 3-4) was collected and the beads were washed extensively before the proteins were eluted (lanes 5-6). Detection of the proteins was by Western blotting (WB) using a polyclonal GST antibody and a monoclonal Flag antibody. The position of the molecular weight markers is indicated on the left in kilodaltons (kDa).

Pab2 is required for Rrp6-dependent processing, but functions in a pathway distinct from the core exosome

To further characterize the functional relationship between Pab2, Rrp6 and the core exosome, double mutants were generated. Deletion of *pab2* from a strain that expressed a catalytically impaired version of Dis3 (*dis3-54*) exacerbated the growth defect of the *dis3-54* single mutant strain at all tested temperatures (Fig. 5A). In contrast, the *pab2Δ rrp6Δ* double mutant strain showed a comparable growth rate to that of the *rrp6Δ* single mutant. These results suggest that whereas Pab2 and Rrp6 function in the same pathway, Pab2 and Dis3 function in distinct pathways. This conclusion was supported by RNA analyses, which demonstrated that the level of 3'-extended polyadenylated forms of snoRNA snR99 as well as the reduction in mature snR99 in the *pab2Δ rrp6Δ* double mutant were similar to either single mutant strains (Fig. 5B, compare lanes 11-12 to lanes 9-10 and 3-4; Fig. 5C). Conversely, the level of 3'-extended poly(A)⁺ forms of snR99 in the *pab2Δ dis3-54* double mutant was markedly increased relative to either single mutant strains (compare lanes 15-16 to 13-14 and 3-4); this was particularly noticeable for 3'-extended species of snR99 with

short poly(A) tails. Furthermore, combining the *dis3-54* allele with the deletion of *pab2* restored the reduction of mature snoRNA detected in the *pab2* Δ single mutant (Fig. 5B-5C).

We also tested whether the polyadenylation of 3'-extended snoRNAs that accumulate in *pab2* Δ cells was dependent on the poly(A) polymerase activity of the TRAMP complex. We therefore deleted *cid14*, which encodes the single catalytic subunit of the fission yeast TRAMP complex (Win *et al.*, 2006). As can be seen in Fig. 5B, the length of poly(A) tails of 3'-extended snR99 in a *pab2* Δ *cid14* Δ double mutant strain was comparable to that of the single *pab2* Δ strain (compare lanes 7-8 to 3-4). Consistently, the poly(A) bodies that specifically accumulated in the nucleus of *pab2* Δ cells were still detected in the *pab2* Δ *cid14* Δ strain (Fig. S6). These results suggest that Pab2 functions independently of the poly(A) polymerase activity of the TRAMP complex. Interestingly, the *pab2* Δ *cid14* Δ double mutant strain exhibited synthetic growth defects (Fig. 5A). Moreover, deletion of *cid14* in the *pab2* Δ strain restored the lower levels of mature snoRNA seen in the *pab2* Δ single mutant (Fig. 5A, compare lanes 7-8 to 3-4; Fig. 5C), similar to the *pab2* Δ *dis3-54* double mutant. Consistent with a nuclear-specific role for Pab2 in exosome-mediated snoRNA processing, cells deleted for *ski7*, which encodes a cytoplasmic-specific exosome-associated protein, did not accumulate 3'-extended polyadenylated snoRNA (Fig. S7). Moreover, *ski7* demonstrated no genetic interaction with *pab2* (Fig. 5A).

Taken together, these data suggest that Pab2 is required for Rrp6-dependent processing of 3'-extended polyadenylated snoRNAs. Conversely, although our results support a role for Dis3 and Cid14 in snoRNA metabolism, they appear to function in a pathway that diverges from that of Pab2.

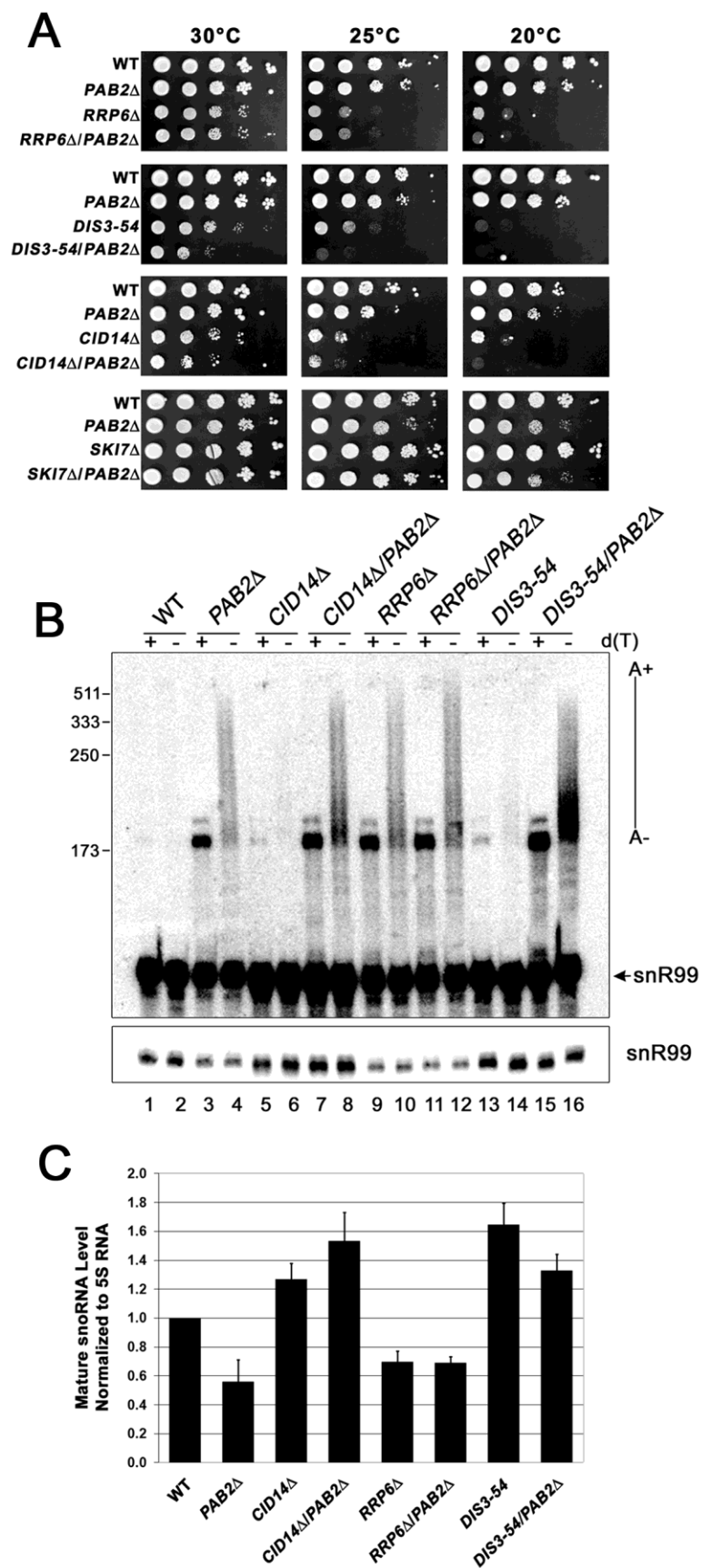


Figure 5. Pab2 is required for Rrp6-dependent processing, but functions in a pathway distinct from the core exosome.

(A) Synthetic growth defects of *pab2Δ dis3-54* and *pab2Δ cid14Δ* double mutants. (B) Equal amounts of total RNA prepared from the indicated strains was treated with RNase H in the presence of a DNA oligonucleotide complementary to H/ACA class snR99. RNase H reactions were performed in the presence (+) or absence (-) of oligo(dT). Size markers (in nucleotides) are indicated on the left. (C) Quantification of alterations in mature snR99 levels, with the wild-type ratio set to 1. Values represent the means of at least three independent experiments and bars correspond to standard deviations.

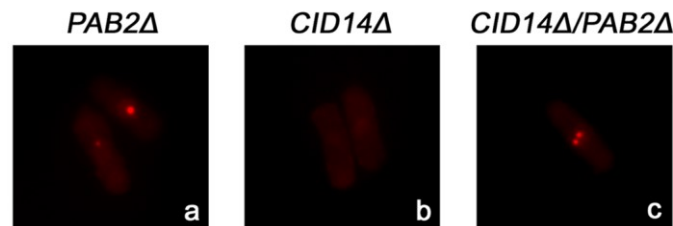


Figure S6. Cid14-independent accumulation of nuclear poly(A) bodies in *pab2Δ* cells. *pab2Δ*, *cid14Δ*, and *pab2Δ cid14Δ* double mutant strains were visualized for polyadenylated RNA using Cy3-labeled oligo(dT)50 (panels a, b, and c, respectively).

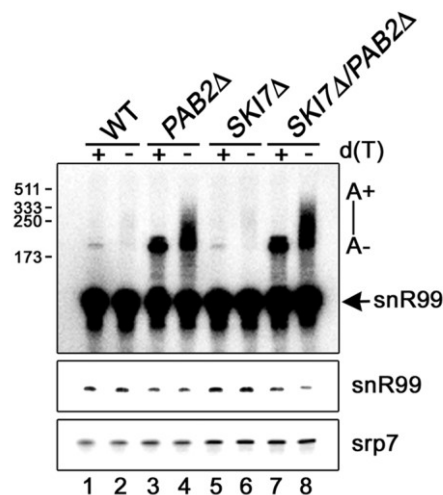


Figure S7. Absence of 3'-extended snoRNA accumulation in *ski7Δ* cells.

Total RNA prepared from the indicated strains was treated with RNase H in the presence of a DNA oligonucleotide complementary to H/ACA class snR99. RNase H reactions were performed in the presence (+) or absence (-) of oligo(dT). Srp7 RNA was used as a loading control. Size markers (in nucleotides) are indicated on the left.

Pab2 is recruited to the 3'-end of snoRNA genes and associates with polyadenylated snoRNAs

mRNA 3'-end processing events are tightly integrated to the transcription cycle through the carboxy-terminal domain of RNA polymerase II (Buratowski, 2005, Hirose et Manley, 2000, Proudfoot, 2004). Accordingly, many factors involved in 3'-end processing/polyadenylation of mRNAs are recruited late during the transcription cycle and near the polyadenylation site of genes (Ahn *et al.*, 2004, Kim *et al.*, 2004). To determine whether Pab2 is associated with sites of snoRNA transcription, ChIP assays were performed using a TAP-tagged version of Pab2 (Lemieux et Bachand, 2009). Two independently-transcribed snoRNA genes were monitored for Pab2 occupancy by ChIP: the C/D class *SNOR68* and the H/ACA class *SNR99*. Pab2 was strongly recruited to both *SNOR68* (Fig. 6A) and *SNR99* (Fig. 6B) genes. Significantly, Pab2 showed the greatest crosslinking signal near the polyadenylation site of both snoRNA genes. These ChIP experiments indicate that Pab2 is recruited to sites of snoRNA transcription and that Pab2 is enriched near the polyadenylation site of two independent snoRNA genes.

We next examined whether Pab2 is bound to 3'-extended polyadenylated snoRNAs by RNA immunoprecipitation (RNA-IP) assays. Because 3'-extended polyadenylated snoRNAs are hardly detectable in wild-type cells (Fig. 2), we performed these experiments in an *rrp6Δ* genetic background as these RNA species accumulate in the absence of Rrp6 (Fig. 5). As can be seen in Fig. 6C, a large enrichment of polyadenylated snR99 was observed in a Pab2 precipitate as compared to a control purification (upper panel, lanes 3-4). In contrast, the abundant cytoplasmic *srp7* RNA was not enriched in precipitates of Pab2 (middle panel, lanes 3-4). No signal was detected in the absence of reverse transcription, indicating that the observed amplification was not due to the presence of residual DNA in the immunoprecipitates (Fig. 6C, lower panels). Real-time RT-PCR analysis using primer sets to different snoRNAs confirmed the specific enrichment for other polyadenylated snoRNAs in Pab2 precipitates (Fig. 6D). These results show that Pab2 is associated with polyadenylated snoRNAs.

We also examined whether the accumulation of 3'-extended snoRNAs perturbed the subcellular localization of Pab2. We therefore visualized the localization of a GFP-tagged version of Pab2 in wild-type and *rrp6Δ* cells. Whereas Pab2-containing foci were rarely observed in normal cells (Fig. 6E, panel b), the absence of Rrp6 caused the concentration of Pab2 in specific nuclear foci (Fig. 6E, panel d). Notably, the Pab2 foci in *rrp6Δ* cells coincided with foci corresponding to 3'-extended forms of snR99 (Fig. 6F, panels a-c). Taken together, these results suggest that poly(A)-bound Pab2 is found at sites of polyadenylated snoRNA accumulation.

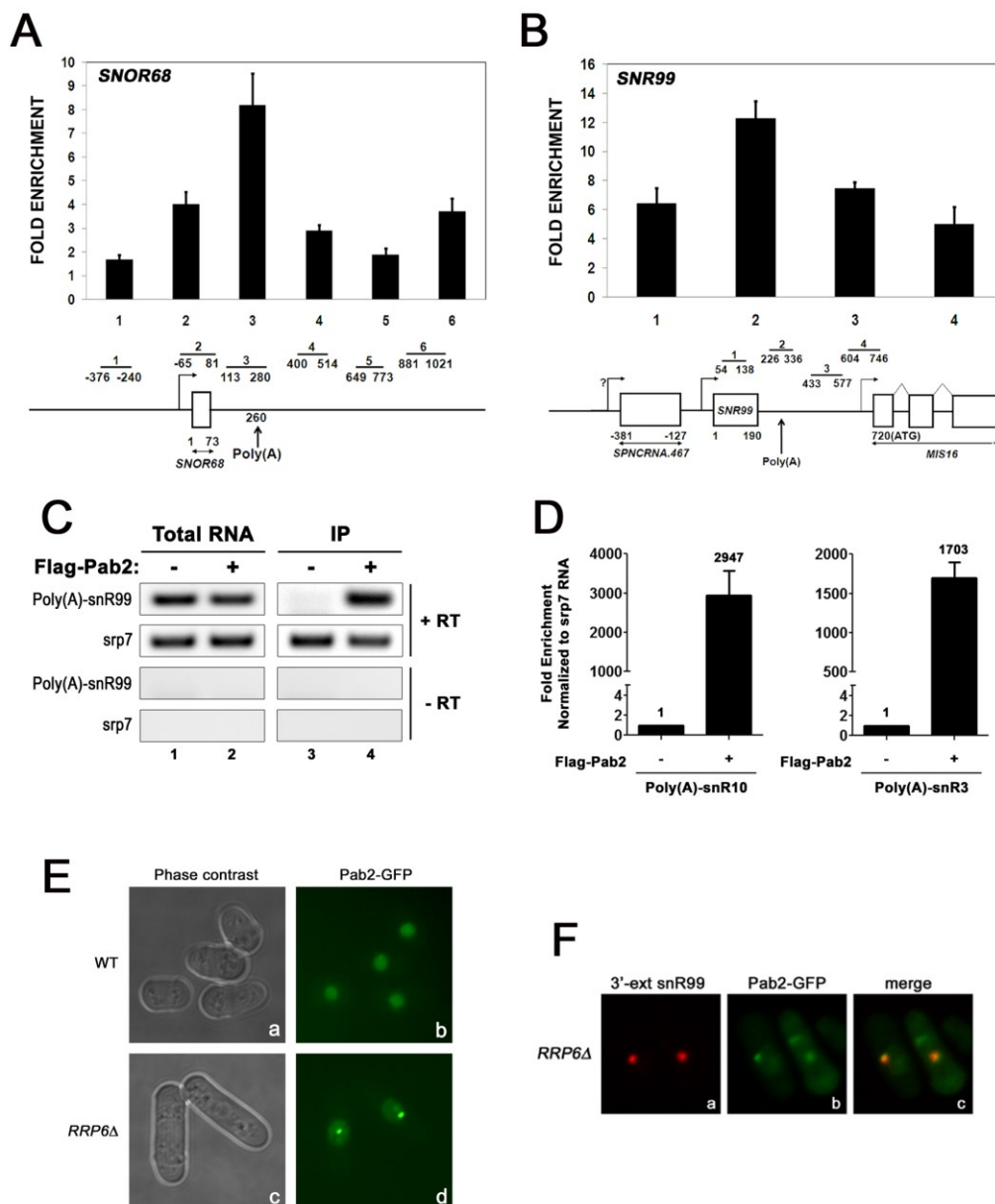


Figure 6. Recruitment of Pab2 to the 3'-end of snoRNA genes.

ChIP assays were performed using a TAP-tagged version of Pab2. Coprecipitating DNA was quantified by real-time PCR using gene-specific primer pairs located along the (A) *snoR68* and (B) *snR99* genes. ChIP data are presented as the fold enrichment of Pab2 relative to a control region. Values represent the means of at least three independent experiments and bars correspond to standard deviations. The position of the polyadenylation sites are indicated (poly(A)). (C) RNA-IP experiments showing that polyadenylated snR99, but not the cytoplasmic *srp7* RNA, is enriched in Flag-Pab2 precipitates. cDNA synthesis for snR99 and *srp7* were oligo d(T)- and random-primed, respectively. (D) Polyadenylated snoRNA association (IP:IN ratio) is normalized to *srp7* RNA. Values were then set to 1.0 for the control immunoprecipitations. Values represent the means of at least two independent experiments and bars correspond to standard deviations. (E) Wild-type (panels a-b) and *rrp6Δ* cells (panels c-d) expressing a GFP-tagged version of Pab2 were visualized by phase

contrast and for Pab2-GFP by live microscopy. (F) *rrp6Δ* cells that expressed GFP-Pab2 were fixed and visualized by FISH using a probe for 3'-extended snR99 (panel a) as well as for GFP fluorescence (panel b).

DISCUSSION

Poly(A)-binding proteins (PAPBs) play essential roles in eukaryotic gene expression. Whereas PABPs play key functions in gene expression via the control of mRNAs, the role of poly(A)-binding proteins in the expression of noncoding RNAs has never been reported. In the course of studying the function of the fission yeast PABPN1 homolog Pab2, we have performed the first genome-wide study that addresses the functional significance of the nuclear poly(A)-binding protein. In addition to identifying a novel function for the nuclear poly(A)-binding protein in the synthesis of noncoding RNAs, our results provide important insights into exosome recruitment to polyadenylated RNA substrates.

Gene-specific regulation by Pab2 in fission yeast

Previous studies about PABPN1, which were based mainly on approaches using purified proteins and *in vitro* polyadenylation assays, have led to a model in which PABPN1 is critical for mRNA polyadenylation (Bienroth *et al.*, 1993, Kerwitz *et al.*, 2003, Wahle, 1991). It was surprising, therefore, given the important role of the 3' poly(A) tail in mRNA expression, that the abundance of most transcripts was unaffected by the deletion of fission yeast *pab2*, as revealed by our microarray experiments. Furthermore, 3'-end analysis of several mRNAs by RNase H mapping between wild-type and *pab2Δ* strains showed no alteration in 3'-end decision and poly(A) tail length (Fig. 2 and data not shown). Our findings argue that Pab2 is not a general factor required for mRNA polyadenylation, and therefore suggest that the regulation of mRNA poly(A) tail synthesis in the nucleus is likely to be more complex than previously anticipated.

The genome-wide approach used in this study indicated that several small nucleolar RNAs (snoRNAs) were upregulated in the absence of Pab2. Although the microarray results

identified mostly H/ACA box snoRNAs (Fig. 1), we showed that the absence of Pab2 results in the accumulation of 3'-extended polyadenylated forms of both C/D and H/ACA class snoRNAs (Fig. 2). This bias toward the detection of H/ACA class snoRNAs in the microarray experiments is likely due to the generally larger size of H/ACA versus C/D box snoRNAs, which improves stable hybridization during microarray experiments. In addition, since RNase H assays detected 3'-extended polyadenylated forms of snoRNAs for which snoRNA probes did not produce usable signals on DNA microarrays, we believe that the number of snoRNAs that were affected by Pab2 was underestimated in our study. Given that poly(A) tails in *S. pombe* are on average 40 nt (Lackner *et al.*, 2007), the accumulation of snoRNAs with poly(A) tails up to 300 nt in the absence of Pab2 (Figs. 2 and 5) explains, at least in part, the hyperadenylation phenotype of *pab2* Δ cells (Perreault *et al.*, 2007).

A function for Pab2 in the synthesis of noncoding RNAs

Given the established role of PABPs in mRNA expression, the accumulation of polyadenylated forms of noncoding snoRNAs in *pab2* Δ cells was surprising. In the past few years, however, it has been established that polyadenylation not only provides stability to RNA molecules, but also contributes to RNA processing via the exosome complex of 3'→5' exonucleases. Accordingly, our study identified the nuclear exosome as the machinery that functions with Pab2 to promote processing of polyadenylated snoRNAs. The direct role of Pab2 in exosome-mediated processing of polyadenylated snoRNAs is supported by several observations: (i) Pab2 is physically associated with Rrp6 and Dis3 *in vivo* (Fig. 4C), two exonucleases specific to the exosome, (ii) Pab2 interacts directly with Rrp6 (Fig. 4D), (iii) microarray data indicate that the absence of Pab2 does not affect the expression of *rrp6* or any of the genes encoding for components of the core exosome (Fig. 1A and data not shown), (iv) Pab2 is recruited to the polyadenylation site of snoRNA genes (Fig. 6), and (v) Pab2 is bound to 3'-extended polyadenylated snoRNAs (Fig. 6). The role of the exosome in the processing of polyadenylated snoRNAs is therefore conserved between fission and budding yeasts as polyadenylated snoRNAs also accumulate in strains of *S. cerevisiae* lacking Rrp6 (Grzechnik et Kufel, 2008, van Hoof *et al.*, 2000, Wyers *et al.*, 2005). Our study therefore provides significant new insights into exosome-dependent processing of polyadenylated

snoRNAs by demonstrating the key role of a poly(A)-binding protein in this process. Interestingly, the genome of *S. cerevisiae* does not encode for a homolog of fission yeast Pab2 and mammalian PABPN1 (Winstall *et al.*, 2000), suggesting that a yet to be identified poly(A)-binding protein is likely involved in exosome-mediated processing of polyadenylated snoRNAs in budding yeast.

Our results also indicated that polyadenylation of the 3'-extended form of a snoRNA accumulating in the *pab2* Δ mutant was independent of the poly(A) polymerase activity of the TRAMP complex. Given that we identified the canonical nuclear poly(A) polymerase, the product of the *S. pombe* *plal* gene (homolog of *S. cerevisiae* *pap1*), as a Pab2-associated protein by tandem affinity purification (Lemieux et Bachand, 2009), we propose that Pla1 is responsible for the polyadenylation of 3'-extended snoRNA that accumulate in *pab2* Δ cells. Consistently, Pap1-dependent polyadenylation of snoRNAs has been reported (Carneiro *et al.*, 2007, Grzechnik et Kufel, 2008, van Hoof *et al.*, 2000, Wyers *et al.*, 2005).

Despite the functional overlap between Rrp6 and Dis3, several studies using budding yeast demonstrate that depletion of these two nucleases can result in different RNA processing defects (Allmang *et al.*, 1999, Dziembowski *et al.*, 2007, Mitchell *et al.*, 2003, van Hoof *et al.*, 2000), suggesting that Rrp6 and Dis3 are responsible for distinctive steps in RNA processing. Furthermore, recent results now show that Rrp6 can perform specific processing events independently of the core exosome (Callahan et Butler, 2008). Our data using fission yeast are consistent with these aforementioned observations: although our genomic data indicated that Dis3, Rrp6, and Pab2 are functionally related (Figs. 4 and S4), the RNA processing defects detected in the *rrp6* Δ and *dis3-54* strains were clearly distinct (Fig. 5). Notably, the defects in snoRNA processing in the *pab2* Δ strain were most similar to those detected in *rrp6* Δ and *pab2* Δ *rrp6* Δ double mutants (Fig. 5). In contrast, the absence of Pab2 expression in the context of a catalytically impaired core exosome (*dis3-54* allele) resulted in levels of polyadenylated snoRNAs greater than either the *pab2* Δ and *dis3-54* single mutant. These results suggest that whereas Pab2 and Rrp6 function in the same pathway, Pab2 and the core exosome participate in overlapping, but distinct pathways associated to snoRNA metabolism. This conclusion is consistent with the greater enrichment

of Pab2 in the Rrp6 purification relative to the Dis3 purification (Fig. 4C) and the direct interaction between Pab2 and Rrp6 (Fig. 4D).

How does Pab2 function in snoRNA synthesis? Although a polyadenylation-independent pathway had previously been proposed to mediate 3'-end formation of snoRNAs in *S. cerevisiae* (Carroll *et al.*, 2004, Fatica *et al.*, 2000, Kim *et al.*, 2006, Steinmetz *et al.*, 2001, Vasiljeva et Buratowski, 2006), recent findings suggest that polyadenylation is involved in 3'-end maturation of snoRNA precursors by the exosome (Grzechnik et Kufel, 2008). Accordingly, the accumulation of 3'-extended polyadenylated snoRNAs together with the concurrent reduction of mature snoRNAs in the *pab2* Δ , *rrp6* Δ , and *pab2* Δ *rrp6* Δ double mutant strains (Figs. 2 and 5) is consistent with a precursor-product relationship. Our ability to copurify polyadenylated snoRNAs with a snoRNP protein (Fig. S8) is also consistent with the polyadenylated species being precursors to mature snoRNAs.

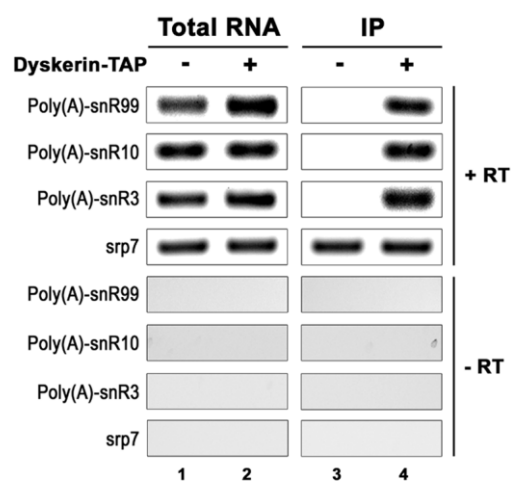


Figure S8. The H/ACA snoRNP protein dyskerin associates with polyadenylated H/ACA snoRNAs.

RT-PCR analysis from total RNA (lanes 1-2) and RNA extracted from IgG-sepharose beads (lanes 3-4) that were prepared from *pab2* Δ cells (lanes 1 and 3) and from *pab2* Δ cells that expressed a TAP-tagged version of dyskerin (lanes 2 and 4). cDNA synthesis of snR99, snR3, and snR10 was oligo d(T)-primed, whereas *srp7* cDNA was random-primed (+RT panel).

These results support a model (Fig. 7) in which 3'-extended polyadenylated snoRNAs that accumulate in *pab2* Δ cells correspond to pre-snoRNAs stalled or delayed in 3'-end

processing, but that mature snoRNAs can still be synthesized via an independent pathway. This independent pathway could involve a fission yeast complex analogous to the Nrd1-Nab3-Sen1 complex that is involved in snoRNA termination and synthesis in budding yeast (Carroll *et al.*, 2004, Kim *et al.*, 2006, Steinmetz *et al.*, 2001, Vasiljeva et Buratowski, 2006). Transcription complexes that escape the first termination pathway reach alternative termination sites that are recognized by an mRNA-like cleavage and polyadenylation machinery (Dheur *et al.*, 2003, Garas *et al.*, 2008). Polyadenylation of released 3'-extended snoRNAs provides high-affinity binding sites for Pab2 and favors the rapid transfer of Pab2 to the growing poly(A) tail. The exonucleolytic activity of Rrp6 carries out the processing of 3'-extended polyadenylated snoRNAs until it reaches a stable snoRNP complex, thus generating the mature snoRNA 3' end. Alternatively, defective snoRNP particles are recognized by exosome/TRAMP and directed to the degradation pathway. Defective particles can nevertheless return to the processing pathway if allowed sufficient time to remodel into stable snoRNPs, as suggested by our results in which the reduced levels of mature snoRNAs seen in *pab2Δ* cells were restored in the *pab2Δ dis3-54* and *pab2Δ cid14Δ* double mutant strains (Fig. 5). Our model in which Pab2 promotes Rrp6 recruitment to poly(A) tails is interesting in light that recombinant Rrp6 is relatively inefficient at degrading a poly(A) substrate *in vitro* (Liu *et al.*, 2006). Together, the aforementioned activities proposed for Pab2 significantly contribute to the efficient processing of polyadenylated snoRNAs since these RNA species are hardly detectable in normal cells.

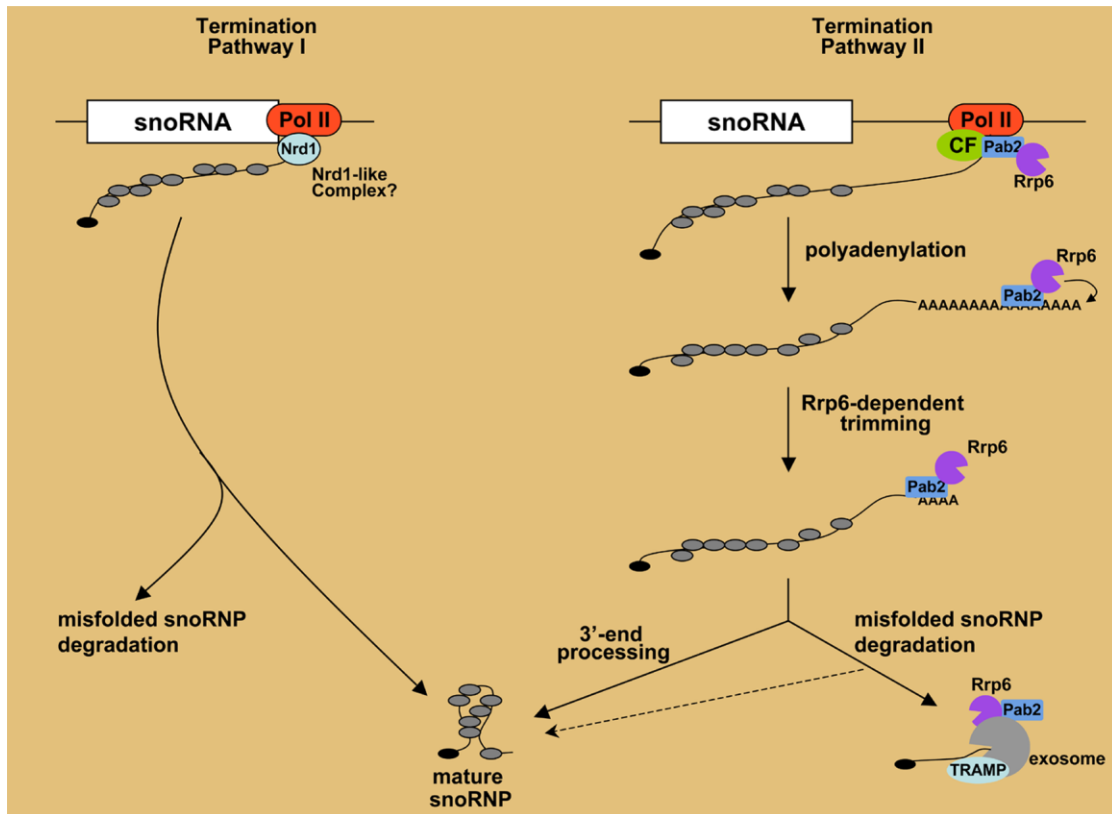


Figure 7. Model for the role of Pab2 in snoRNA synthesis.

snoRNP proteins are cotranscriptionally recruited to nascent snoRNAs to control the biogenesis of snoRNPs. Two independent pathways function in snoRNA synthesis. The first generates mature snoRNPs via a system related to the budding yeast Nrd1/Nab3 complex using termination pathway I. Transcription complexes that escape the first termination system encounter a fail-safe termination pathway that uses an mRNA-like cleavage and polyadenylation complex (CF) to release 3'-extended pre-snoRNAs. Polyadenylation of pre-snoRNAs by the canonical poly(A) polymerase leads to the transfer of a Pab2/Rrp6 complex to the growing poly(A) tail, and efficient 3'-end processing by Rrp6. Properly assembled snoRNPs will be shielded against complete exonucleolytic digestion by Rrp6. Defective snoRNAs that are not assembled into proper particles are directed to degradation by the nuclear exosome, but can produce mature snoRNAs if pre-snoRNPs are allowed sufficient time for remodeling into proper particles.

Our study also provides interesting insights into the subcellular localization where processing of polyadenylated snoRNAs takes place. We found that a 3'-extended polyadenylated snoRNA localized to discrete nucleolar foci in *pab2Δ* and *rrp6Δ* cells. This echoes a report that polyadenylated snoRNAs accumulate in a specific nucleolar domain in *rrp6Δ* mutant of budding yeast (Carneiro *et al.*, 2007). We further showed that Pab2 and a 3'-extended snoRNA were colocalized in similar foci in the *rrp6Δ* strain, suggesting that

these foci represent processing centers where maturation of polyadenylated snoRNAs is delayed due to the absence of Rrp6. We envisage that misassembled snoRNPs that are directed to the exosome/TRAMP discard pathway are also degraded in these foci.

Conclusions

Our findings unveiled a previously uncharacterized function for a poly(A)-binding protein in the expression of noncoding RNAs. Given the extensive similarity between the yeast and human nuclear poly(A)-binding proteins, it is likely that the function of Pab2 in exosome-dependent RNA processing described here is evolutionarily conserved. The detection of poly(A) RNA-containing aggregates in the nucleus of muscle cells of individuals with mutations in *PABPN1* is a pathophysiological hallmark of the neuromuscular disorder, oculopharyngeal muscular dystrophy (OPMD). The intense accumulation of polyadenylated RNAs in a specific region of the nucleus brought about by the absence of *S. pombe* Pab2 is striking and raises the possibility that the intranuclear aggregates in OPMD are related to defects in the processing of noncoding RNAs.

EXPERIMENTAL PROCEDURES

DNA microarray analysis

cDNA synthesis of total RNAs, labeling, and microarray hybridization procedures have been described previously (Bachand *et al.*, 2006). We determined statistical significance using SAM (Significance Analysis of Microarrays) (Tusher *et al.*, 2001). Briefly, SAM provides a statistical value calculated for each gene based on the change in gene expression relative to the standard deviation of repeated measurements. The false discovery rate was set to be below 5%. The microarray data were deposited into the ArrayExpress database and the accession number is E-TABM-843.

RNase H analysis

Total yeast RNA was RNase H-treated in a mixture containing message-specific oligonucleotides plus or minus oligo(dT) as previously described (Decker et Parker, 1993). RNA samples were resolved on 6% polyacrylamide-8 M urea gel, transferred onto nylon membranes, and probed using [³²P]-labeled gene-specific probes.

Fluorescent in situ hybridization (FISH)

3'-extended snR99 and poly(A)+ RNA were visualized by FISH in *S. pombe* using a previously described method (Perreault *et al.*, 2008). The 54-mer complementary to 3'-extended snR99 was conjugated with Cy5. The Cy5 signal in Fig. 3A was converted to blue color using Adobe Photoshop to distinguish Cy5 and Cy3 signals. Control experiments using Cy5- and Cy3-labeled probes individually on *pab2Δ* cells confirmed the absence of bleed-through signal in the other channels (data not shown).

Expression of recombinant proteins and in vitro pull-down assays

GST and GST-Pab2 were expressed in *E. coli* and purified as described previously (Perreault *et al.*, 2007). *S. pombe* Rrp6 was amplified by PCR using a fission yeast cDNA library (a generous gift from Charlie Hoffman) and expressed with a C-terminal Flag epitope in *E. coli*. For *in vitro* pull-down assays, equal amounts of Rrp6-coated beads were incubated with 1 mg of GST or GST-Pab2. The bound proteins were analyzed by immunoblotting using an Odyssey infrared imaging system.

RNA immunoprecipitation experiments

RNA IPs were performed as described previously (Ferreira-Cerca *et al.*, 2007, Tardiff *et al.*, 2006) using extracts prepared from a *pab2Δ rrp6Δ* double mutant strain that was previously transformed with a plasmid that express a Flag-tagged version of Pab2 or an

empty control vector. Briefly, whole-cell extract was incubated with anti-Flag M2 beads. Following extensive washing steps, RNA was extracted from beads (IP RNA) and whole-cell extract (input RNA) using hot-phenol. RNA samples were treated with DNase I (Promega) and cDNA synthesis was primed with random hexamers (for *srp7* RNA) and oligo d(T) (for polyadenylated snoRNAs). Quantitative real-time PCR was performed on a Realplex PCR machine (Eppendorf). Reported values are the averages of at least two independent experiments, and error bars represent ± 1 SD.

ACKNOWLEDGMENTS

We thank R. Maraia, M. Yaganida, and P. Beauregard for strains and plasmids; P. Chartrand and F. Gallardo for advice on FISH microscopy; S. Abou Elela and F. Dragon for critical reading of the manuscript. We also thank S. Khadayate for microarray data deposition. This work was funded by grants from NSERC to F. B. and from Cancer Research UK to J.B. J.F.L. was supported by a Scholarship from NSERC. F.B. is the recipient of a New Investigator Award from the CIHR.

REFERENCES

1. Ahn, S. H., Kim, M., and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* 13, 67-76.
2. Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999a). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J* 18, 5399-5410.
3. Allmang, C., Mitchell, P., Petfalski, E., and Tollervey, D. (2000). Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res* 28, 1684-1691.

4. Allmang, C., Petfalski, E., Podtelejnikov, A., Mann, M., Tollervey, D., and Mitchell, P. (1999b). The yeast exosome and human PM-Scl are related complexes of 3'→5' exonucleases. *Genes Dev* 13, 2148-2158.
5. Arigo, J. T., Eyler, D. E., Carroll, K. L., and Corden, J. L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell* 23, 841-851.
6. Bachand, F., Lackner, D. H., Bahler, J., and Silver, P. A. (2006). Autoregulation of ribosome biosynthesis by a translational response in fission yeast. *Mol Cell Biol* 26, 1731-1742.
7. Bienroth, S., Keller, W., and Wahle, E. (1993). Assembly of a processive messenger RNA polyadenylation complex. *EMBO J* 12, 585-594.
8. Bousquet-Antonelli, C., Presutti, C., and Tollervey, D. (2000). Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* 102, 765-775.
9. Buratowski, S. (2005). Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol* 17, 257-261.
10. Callahan, K. P., and Butler, J. S. (2008). Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p. *Nucleic Acids Res* 36, 6645-6655.
11. Carneiro, T., Carvalho, C., Braga, J., Rino, J., Milligan, L., Tollervey, D., and Carmo-Fonseca, M. (2007). Depletion of the yeast nuclear exosome subunit Rrp6 results in accumulation of polyadenylated RNAs in a discrete domain within the nucleolus. *Mol Cell Biol* 27, 4157-4165.
12. Carroll, K. L., Pradhan, D. A., Granek, J. A., Clarke, N. D., and Corden, J. L. (2004). Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol Cell Biol* 24, 6241-6252.
13. Chekanova, J. A., Gregory, B. D., Reverdatto, S. V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., et al. (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell* 131, 1340-1353.
14. Decker, C. J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev* 7, 1632-1643.

15. Dez, C., Houseley, J., and Tollervey, D. (2006). Surveillance of nuclear-restricted pre-ribosomes within a subnucleolar region of *Saccharomyces cerevisiae*. *EMBO J* 25, 1534-1546.
16. Dheur, S., Voile, T. A., Voisinet-Hakil, F., Minet, M., Schmitter, J. M., Lacroute, F., Wyers, F., and Minvielle-Sebastia, L. (2003). Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. *EMBO J* 22, 2831-2840.
17. Dziembowski, A., Lorentzen, E., Conti, E., and Seraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol* 14, 15-22.
18. Fatica, A., Morlando, M., and Bozzoni, I. (2000). Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *EMBO J* 19, 6218-6229.
19. Ferreira-Cerca, S., Poll, G., Kuhn, H., Neueder, A., Jakob, S., Tschochner, H., and Milkereit, P. (2007). Analysis of the in vivo assembly pathway of eukaryotic 40S ribosomal proteins. *Mol Cell* 28, 446-457.
20. Garas, M., Dichtl, B., and Keller, W. (2008). The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *RNA* 14, 2671-2684.
21. Grzechnik, P., and Kufel, J. (2008). Polyadenylation linked to transcription termination directs the processing of snoRNA precursors in yeast. *Mol Cell* 32, 247-258.
22. Hirose, Y., and Manley, J. L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev* 14, 1415-1429.
23. Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* 7, 529-539.
24. Houseley, J., and Tollervey, D. (2006). Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep* 7, 205-211.
25. Ibrahim, H., Wilusz, J., and Wilusz, C. J. (2008). RNA recognition by 3'-to-5' exonucleases: the substrate perspective. *Biochim Biophys Acta* 1779, 256-265.
26. Kadaba, S., Wang, X., and Anderson, J. T. (2006). Nuclear RNA surveillance in *Saccharomyces cerevisiae*: Trf4p-dependent polyadenylation of nascent hypomethylated tRNA and an aberrant form of 5S rRNA. *RNA* 12, 508-521.

27. Kerwitz, Y., Kuhn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E., and Wahle, E. (2003). Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *EMBO J* 22, 3705-3714.
28. Kim, M., Ahn, S. H., Krogan, N. J., Greenblatt, J. F., and Buratowski, S. (2004). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J* 23, 354-364.
29. Kim, M., Vasiljeva, L., Rando, O. J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct pathways for snoRNA and mRNA termination. *Mol Cell* 24, 723-734.
30. Kiss, T., Fayet, E., Jady, B. E., Richard, P., and Weber, M. (2006). Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb Symp Quant Biol* 71, 407-417.
31. Kuhn, U., and Wahle, E. (2004). Structure and function of poly(A) binding proteins. *Biochim Biophys Acta* 1678, 67-84.
32. LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* 121, 713-724.
33. Lackner, D. H., Beilharz, T. H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Preiss, T., and Bahler, J. (2007). A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol Cell* 26, 145-155.
34. Lebreton, A., and Seraphin, B. (2008). Exosome-mediated quality control: substrate recruitment and molecular activity. *Biochim Biophys Acta* 1779, 558-565.
35. Lemieux, C., and Bachand, F. (2009). Cotranscriptional recruitment of the nuclear poly(A)-binding protein Pab2 to nascent transcripts and association with translating mRNPs. *Nucleic Acids Res* 37, 3418-3430.
36. Liu, Q., Greimann, J. C., and Lima, C. D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* 127, 1223-1237.
37. Mangus, D. A., Evans, M. C., and Jacobson, A. (2003). Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* 4, 223.

38. Milligan, L., Torchet, C., Allmang, C., Shipman, T., and Tollervey, D. (2005). A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Mol Cell Biol* 25, 9996-10004.
39. Mitchell, P., Petfalski, E., Houalla, R., Podtelejnikov, A., Mann, M., and Tollervey, D. (2003). Rrp47p is an exosome-associated protein required for the 3' processing of stable RNAs. *Mol Cell Biol* 23, 6982-6992.
40. Murakami, H., Goto, D. B., Toda, T., Chen, E. S., Grewal, S. I., Martienssen, R. A., and Yanagida, M. (2007). Ribonuclease activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. *PLoS ONE* 2, e317.
41. Perreault, A., Bellemer, C., and Bachand, F. (2008). Nuclear export competence of pre-40S subunits in fission yeast requires the ribosomal protein Rps2. *Nucleic Acids Res* 36, 6132-6142.
42. Perreault, A., Lemieux, C., and Bachand, F. (2007). Regulation of the Nuclear Poly(A)-binding Protein by Arginine Methylation in Fission Yeast. *J Biol Chem* 282, 7552-7562.
43. Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* 16, 272-278.
44. Reichow, S. L., Hamma, T., Ferre-D'Amare, A. R., and Varani, G. (2007). The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res* 35, 1452-1464.
45. Saguez, C., Schmid, M., Olesen, J. R., Ghazy, M. A., Qu, X., Poulsen, M. B., Nasser, T., Moore, C., and Jensen, T. H. (2008). Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. *Mol Cell* 31, 91-103.
46. Schmid, M., and Jensen, T. H. (2008). The exosome: a multipurpose RNA-decay machine. *Trends Biochem Sci* 33, 501-510.
47. Steinmetz, E. J., Conrad, N. K., Brow, D. A., and Corden, J. L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413, 327-331.
48. Steinmetz, E. J., Warren, C. L., Kuehner, J. N., Panbehi, B., Ansari, A. Z., and Brow, D. A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell* 24, 735-746.

49. Tardiff, D. F., Lacadie, S. A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol Cell* 24, 917-929.
50. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the *nrd1-nab3* pathway in genome surveillance. *Mol Cell* 23, 853-864.
51. Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98, 5116-5121.
52. van Hoof, A., Lennertz, P., and Parker, R. (2000). Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol* 20, 441-452.
53. Vanacova, S., and Stefl, R. (2007). The exosome and RNA quality control in the nucleus. *EMBO Rep* 8, 651-657.
54. Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol* 3, e189.
55. Vasiljeva, L., and Buratowski, S. (2006). Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell* 21, 239-248.
56. Vasiljeva, L., Kim, M., Terzi, N., Soares, L. M., and Buratowski, S. (2008). Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. *Mol Cell* 29, 313-323.
57. Wahle, E. (1991). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66, 759-768.
58. Wahle, E. (1995). Poly(A) tail length control is caused by termination of processive synthesis. *J Biol Chem* 270, 2800-2808.
59. Wang, S. W., Stevenson, A. L., Kearsey, S. E., Watt, S., and Bahler, J. (2008). Global role for polyadenylation-assisted nuclear RNA degradation in posttranscriptional gene silencing. *Mol Cell Biol* 28, 656-665.
60. West, S., Gromak, N., Norbury, C. J., and Proudfoot, N. J. (2006). Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Mol Cell* 21, 437-443.

61. Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., and Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 1239-1243.
62. Win, T. Z., Draper, S., Read, R. L., Pearce, J., Norbury, C. J., and Wang, S. W. (2006). Requirement of fission yeast Cid14 in polyadenylation of rRNAs. *Mol Cell Biol* 26, 1710-1721.
63. Winstall, E., Sadowski, M., Kuhn, U., Wahle, E., and Sachs, A. B. (2000). The *Saccharomyces cerevisiae* RNA-binding protein Rbp29 functions in cytoplasmic mRNA metabolism. *J Biol Chem* 275, 21817-21826.

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

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

Media and growth conditions

Strains of *S. pombe* used in this study are provided in Table S2. *S. pombe* was grown at 30°C in yeast extract medium with amino acid supplements (YES). Cells were transformed with PCR products by the lithium acetate method. All gene disruptions were performed by PCR-mediated gene targeting (Bahler et al., 1998) using 100-nt oligonucleotides with 80-nt from the appropriate regions of the genomic sequences to be targeted. The oligonucleotide sequences used for the construction of these strains are available upon request. All gene knockouts were confirmed by RT-PCR.

DNA microarray analysis

cDNA synthesis of total RNAs, labeling, and microarray hybridization procedures have been described previously (Bachand et al., 2006). Total RNA was obtained from early to mid-log phase cells and was prepared as previously described (Bachand et al., 2006). Results from microarray hybridization were analyzed using Genepix (Axon instruments) and GeneSpring (Agilent) softwares and the raw data was filtered and normalized as previously described (Lyne et al., 2003). Microarray analyses of total RNAs were performed from at least three biological replicates. According to the microarray hybridization signals, 4,724 genes were detected in at least 2 of the 3 biological replicates for RNAs prepared from wild-type and *pab2*Δ cells, including a dye-swap experiment. We determined statistical significance using

SAM (Significance Analysis of Microarrays) (Tusher et al., 2001). Briefly, SAM provides a statistical value calculated for each gene based on the change in gene expression relative to the standard deviation of repeated measurements. Using the gene set detected in at least 2 of the 3 biological repeats between wild-type and *pab2* deletion mutant (n=4,724), SAM performed a one-class comparison and generated a set of 198 significant genes. The false discovery rate was set to be below 5%. Eleven snoRNAs were found to be upregulated ≥ 1.25 -fold in the absence of Pab2 (Fig. 1B). Of those 11 upregulated snoRNAs, 4 were identified by SAM and are included in Table S1. snoRNA.52, 47, 35, 50, 20, 34, and 25 were not identified by SAM, however, because the standard deviation between the three biological replicates was larger than the arbitrary cutoff established by SAM, which is very conservative when only three repeats are considered. This variability in signal intensity using microarrays is not unexpected given the small size of these RNAspecies. The significance of gene list overlaps was calculated using a standard Fisher's exact test, and the p-values were adjusted with a Bonferroni multiple testing correction.

RNase H analysis

Total yeast RNA (between 15-45 μ g) was RNase H-treated in a mixture containing 300 ng of message-specific oligonucleotide plus or minus up to 500 ng of oligo(dT) in buffer

condition as previously described (Decker and Parker, 1993). RNA samples were resolved on 6% polyacrylamide-8 M urea gel in 1X TBE. Following the transfer onto nylon membranes, RNAs were crosslinked using a UV Stratalinker. Membranes were probed using gene-specific oligonucleotides that were 5'-labeled with [^{32}P] using T4 polynucleotide kinase. After extensive washing steps, membranes were exposed to Phosphor Screens and visualized using STORM 860 or Typhoon Trio instruments and analysed using the ImageQuant TL software.

Chromatin immunoprecipitation (ChIP) experiments

ChIPs were performed as previously described (Lemieux and Bachand, 2009) using either a mouse monoclonal antibody (8WG16) specific to the carboxy-terminal domain (CTD) of the large subunit of RNA polymerase II or IgG-sepharose. Quantification of the immunoprecipitated DNA was done by quantitative real-time PCR as described previously (Lemieux and Bachand, 2009). Specific dilutions of coimmunoprecipitated DNA as well as of sonicated and reverse cross-linked input DNA were used to determine the fold enrichment in each immunoprecipitate. To calculate the increase in signal in the different gene regions, the percentage input values obtained by quantitative PCR were normalized to the percentage input value obtained with the nontranscribed intergenic region, arbitrarily set to 1. The 104-bp PCR amplicon corresponding to the intergenic region lays on the left arm of chromosome I (nucleotides 3009380 to 3009484). Gene-specific primer sets are available upon request. Each PCR was run in triplicate, and all chromatin immunoprecipitations were repeated at least two times using independent chromatin extracts.

Fluorescent in situ hybridization (FISH)

3'-extended snR99 and poly(A)+ RNA were visualized by fluorescence in situ hybridization in *S. pombe* using a previously described method (Perreault et al., 2008). A DNA oligonucleotide complementary to 54 bases of the snR99 3' extension, 5'-XCATGTAGTATXGCAAACCTTTGGGXACGTATCTAAACXAATCAAATATAAACX T-3' (where X represent amino-allyl thymine) as well as a Cy3-labeled oligo d(T)50 were

synthesized (Integrated DNA Technologies, Coralville). The 54-mer complementary to 3'-extended snR99 was conjugated with Cy5. The Cy5 signal in Fig. 3A was converted to blue color using Adobe Photoshop to distinguish Cy5 and Cy3 signals. Control experiments using Cy5- and Cy3-labeled probes individually on *pab2* Δ cells confirmed the absence of bleed-through signal in the other channels (data not shown).

Protein analysis

For the affinity purification experiments, ~50-ml of mid-log phase cells were resuspended in ice-cold lysis buffer (20 mM Tris-HCl pH 7.5; 10 mM MgCl₂; 1 mM DTT; 50 mM KCl) containing a cocktail of protease inhibitors, and lysed using glass beads in a Fastprep FP120. Clarified lysates were normalized for total protein concentration and 1 mg of total proteins were subjected to immunoprecipitation using IgG-sepharose. Eluted proteins were separated on 12% SDS-PAGE, transferred to nitrocellulose membranes, and analysed by immunoblotting using a rabbit polyclonal antibody specific to protein A (Sigma) and a mouse monoclonal antibody specific to the HA epitope (Roche). Membranes were then probed with goat anti-rabbit and antimouse secondary antibodies conjugated to IRdye 800 (Rockland Immunochemicals) and AlexaFluor 680 (Molecular Probes), respectively. Linear detection of the proteins was performed and quantified using an Odyssey infrared imaging system (LI-COR).

Expression of recombinant proteins and in vitro pull-down assays

GST and GST-Pab2 were expressed in *E. coli* and purified as described previously (Perreault et al., 2007). *S. pombe* Rrp6 was amplified by PCR using a fission yeast cDNA library (a generous gift from Charlie Hoffman) and expressed with a C-terminal Flag epitope in *E. coli* BL21 DE3 cells using the pET-28b(+) vector. Protein expression was induced by the addition of 0.4 mM IPTG for 18h at 18°C. Following centrifugation of the cells and subsequent washing in ice-cold phosphate buffer saline (PBS), the bacterial cells were broken by sonication and the cell membranes were solubilized by the addition of triton X-100 to a final concentration of 1%. The cell debris was sedimented by a 10-min centrifugation at

12,000 rpm at 4°C, and the clarified lysates were subjected to agarose anti-Flag M2 beads (Sigma). After extensive washing, Rrp6 was either kept on beads or eluted by incubations with 200 ng/μl Flag peptide in Tris-Hcl pH 8.6; 250 mM NaCl; 0.1% triton X-100, 10% glycerol. For *in vitro* pull-down assays, equal amounts of Rrp6-coated beads that had previously been pre-equilibrated in TMN150 buffer (150 mM NaCl, 50 mM Tris (pH 8), 5 mM MgCl₂, 0.1% NP-40, and protease inhibitors) were incubated with 1 μg of GST or GST-Pab2 in TMN150 buffer for 1h at 4°C. The beads were then washed three times with 1 ml of TMN150 buffer and the bound proteins eluted by incubating for 5 min at 95°C in SDS-PAGE sample buffer. The Rrp6-Flag, GST, and GST-Pab2 proteins were analyzed by immunoblotting using an Odyssey infrared imaging system (LI-COR) using a mouse monoclonal anti-Flag (Sigma) and a rabbit polyclonal anti- GST (Sc-459, Santa Cruz Biotech).

RNA immunoprecipitation experiments

RNA immunoprecipitations were performed as described previously (Ferreira-Cerca et al., 2007; Tardiff et al., 2006) using extracts prepared from a *pab2Δ rrp6Δ* double mutant strain that was previously transformed with a plasmid that express a Flag-tagged version of Pab2 or an empty control vector. Briefly, cells lysates were prepared using glass beads in buffer A200 containing 200 mM KCl, 20 mM Tris (pH 8), 5 mM MgAc, 1 mM DTT, as well as protease and RNase inhibitors. Extracts were clarified by two consecutive centrifugations at 14,000 rpm for 5 and 10 min at 4°C. Triton X-100 was added to the supernatants to 0.2%. Whole-cell extract (1 mg) was incubated with 10 μl of agarose anti-Flag M2 beads (Sigma) for 90 min at 4°C in A200T (200 mM KCl, 20 mM Tris [pH 8], 5 mM MgAc, 1 mM DTT, 0.2% Triton X-100, protease and RNase inhibitors). Beads were washed three times with 0.5 ml of buffer A200T and RNA was extracted from beads (IP RNA) and whole-cell extract (input RNA) using hot-phenol. RNA samples were treated with DNase I (Promega) and cDNA synthesis was primed with random hexamers (for *srp7* RNA) and oligo d(T) (for polyadenylated snoRNAs). Quantitative real-time PCR was performed on a Realplex PCR machine (Eppendorf). Reported values are the averages of at least two independent experiments, and error bars represent ±1 SD.

SUPPLEMENTARY REFERENCES**(associated with Supplementary experimental procedures, figures and tables)**

65. Bachand, F., Lackner, D. H., Bahler, J., and Silver, P. A. (2006). Autoregulation of ribosome biosynthesis by a translational response in fission yeast. *Mol Cell Biol* 26, 1731-1742.
66. Bahler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., 3rd, Steever, A. B., Wach, A., Philippsen, P., and Pringle, J. R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* 14, 943-951.
67. Decker, C. J., and Parker, R. (1993). A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev* 7, 1632-1643.
68. Ferreira-Cerca, S., Poll, G., Kuhn, H., Neueder, A., Jakob, S., Tschochner, H., and Milkereit, P. (2007). Analysis of the in vivo assembly pathway of eukaryotic 40S ribosomal proteins. *Mol Cell* 28, 446-457.
69. Huang, Y., Bayfield, M.A., Intine, R.V., and Marais R.J. (2006). Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. *Nat Struct Mol Biol* 13, 611-8
70. Lemieux, C., and Bachand, F. (2009). Cotranscriptional recruitment of the nuclear poly(A)-binding protein Pab2 to nascent transcripts and association with translating mRNPs. *Nucleic Acids Res* 37, 3418-3430.
71. Lyne, R., Burns, G., Mata, J., Penkett, C. J., Rustici, G., Chen, D., Langford, C., Vetrie, D., and Bahler, J. (2003). Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics* 4, 27.
72. Perreault, A., Bellemer, C., and Bachand, F. (2008). Nuclear export competence of pre-40S subunits in fission yeast requires the ribosomal protein Rps2. *Nucleic Acids Res* 36, 6132-6142.
73. Perreault, A., Lemieux, C., and Bachand, F. (2007). Regulation of the Nuclear Poly(A)-binding Protein by Arginine Methylation in Fission Yeast. *J Biol Chem* 282, 7552-7562.

74. Tardiff, D. F., Lacadie, S. A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol Cell* 24, 917-929.
75. Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98, 5116-5121.

SUPPLEMENTARY TABLES

Supplementary Table S1. List of up- and down-regulated genes in *pab2Δ* cells

Up-regulated genes		Down-regulated genes	
Systematic ID	Average ratio	Systematic ID	Average ratio
<i>SPCC1223.12c</i>	5,04	<i>SPBC16E9.12c</i>	0,06
<i>SPSNORNA.51</i>	4,9	<i>SPAC3G6.07</i>	0,08
<i>SPBP8B7.04</i>	4,48	<i>SPAC13F5.03c</i>	0,1
<i>SPAC1250.05</i>	4,45	<i>SPBC1198.14c</i>	0,12
<i>SPAC1039.02</i>	4,21	<i>SPAC23H3.15c</i>	0,12
<i>SPBC646.17c</i>	4,15	<i>SPCC737.04</i>	0,14
<i>SPAP27G11.08c</i>	4,12	<i>SPCPB16A4.06c</i>	0,15
<i>SPAC1556.06</i>	4,09	<i>SPAPB1A11.02</i>	0,16
<i>SPAC29E6.07</i>	3,91	<i>SPBC16E9.16c</i>	0,16
<i>SPSNORNA.32</i>	3,66	<i>SPAC869.09</i>	0,17
<i>SPAC25G10.04c</i>	3,52	<i>SPBP4H10.09</i>	0,17
<i>SPCC1393.07c</i>	3,43	<i>SPBC2G2.17c</i>	0,18
<i>SPBC1711.15c</i>	3,38	<i>SPBC2F12.09c</i>	0,18
<i>SPBPB2B2.06c</i>	3,21	<i>SPAC3G9.11c</i>	0,18
<i>SPAC2G11.05c</i>	2,96	<i>SPCC338.18</i>	0,22
<i>SPBC1711.02</i>	2,93	<i>SPAC1751.01c</i>	0,23
<i>SPBPB2B2.03c</i>	2,92	<i>SPBC3E7.02c</i>	0,26
<i>SPAC14C4.03</i>	2,84	<i>SPBC1198.01</i>	0,27
<i>SPBC29A10.02</i>	2,84	<i>SPBP4G3.03</i>	0,27
<i>SPAC23C4.07</i>	2,81	<i>SPNCRNA.32</i>	0,28
<i>SPCC1235.01</i>	2,65	<i>SPBP4H10.10</i>	0,28
<i>SPBC582.06c</i>	2,54	<i>SPAC869.08</i>	0,28
<i>SPAC57A10.04</i>	2,48	<i>SPBC56F2.06</i>	0,29
<i>SPAC23C11.10</i>	2,41	<i>SPAC15E1.02c</i>	0,3
<i>SPAC6C3.05</i>	2,41	<i>SPAC23C11.06c</i>	0,3
<i>SPCC11E10.03</i>	2,4	<i>SPAC32A11.02c</i>	0,31
<i>SPAC27D7.13c</i>	2,38	<i>SPCC16A11.15c</i>	0,32
<i>SPAC222.15</i>	2,37	<i>SPCC794.04c</i>	0,32
<i>SPAC13G7.13c</i>	2,33	<i>SPBPB21E7.01c</i>	0,33

<i>SPCC569.07</i>	2,32	<i>SPBC19C7.04c</i>	0,33
<i>SPSNORNA.37</i>	2,23	<i>SPBC1105.14</i>	0,34
<i>SPBC16C6.10</i>	2,18	<i>SPCC191.01</i>	0,34
<i>SPBC428.07</i>	2,17	<i>SPAC343.12</i>	0,35
<i>SPSNORNA.40</i>	2,12	<i>SPBC359.04c</i>	0,35
<i>SPAC2E1P3.05c</i>	2,04	<i>SPAC16A10.01</i>	0,36
<i>SPAC17A5.18c</i>	2,02	<i>SPCC4G3.03</i>	0,36
<i>SPCC11E10.05c</i>	1,97	<i>SPAC19D5.01</i>	0,37
<i>SPAC22F3.03c</i>	1,95	<i>SPBPB21E7.04c</i>	0,37
<i>SPAC13D1.01c</i>	1,94	<i>SPAC2H10.01</i>	0,38
<i>SPAC1006.03c</i>	1,93	<i>SPAC22H12.01c</i>	0,4
<i>SPBC1E8.04</i>	1,92	<i>SPBC725.10</i>	0,4
<i>SPAC32A11.01</i>	1,92	<i>SPCC1322.08</i>	0,41
<i>SPCC4B3.10c</i>	1,89	<i>SPBPB10D8.01</i>	0,41
<i>SPBC9B6.02c</i>	1,89	<i>SPBC23E6.03c</i>	0,43
<i>SPSNORNA.46</i>	1,87	<i>SPBC365.12c</i>	0,43
<i>SPAC9.04</i>	1,86	<i>SPAC26H5.08c</i>	0,44
<i>SPCC18B5.01c</i>	1,85	<i>SPBC359.06</i>	0,45
<i>SPAPB15E9.03c</i>	1,83	<i>SPBC16D10.08c</i>	0,47
<i>SPBP35G2.03c</i>	1,83	<i>SPBP8B7.05c</i>	0,48
<i>SPAC27E2.08</i>	1,81	<i>SPAC26F1.14c</i>	0,48
<i>SPAC1006.04c</i>	1,81	<i>SPCC569.08c</i>	0,49
<i>SPAC19D5.09c</i>	1,81	<i>SPAC22F8.05</i>	0,49
<i>SPNCRNA.123</i>	1,79	<i>SPNCRNA.132</i>	0,5
<i>SPAC167.08</i>	1,78	<i>SPBC23G7.17c</i>	0,5
<i>SPBC28E12.05</i>	1,78	<i>SPAC167.06c</i>	0,5
<i>SPAC513.03</i>	1,74	<i>SPAC8C9.03</i>	0,51
<i>SPBC29A10.14</i>	1,74	<i>SPBC725.03</i>	0,52
<i>SPBC23G7.12c</i>	1,67	<i>SPBC19C2.04c</i>	0,53
<i>SPBC26H8.03</i>	1,59	<i>SPAC1F7.08</i>	0,55
<i>SPCC569.05c</i>	1,57	<i>SPAC1F7.07c</i>	0,55
<i>SPBC18E5.03c</i>	1,57	<i>SPAC17A2.11</i>	0,55
<i>SPCC1682.08c</i>	1,55	<i>SPAC1420.01c</i>	0,55
<i>SPCC736.02</i>	1,53	<i>SPNCRNA.101</i>	0,55
<i>SPAC30C2.03</i>	1,53	<i>SPNCRNA.26</i>	0,56
<i>SPCC1494.11c</i>	1,52	<i>SPAC17A2.09c</i>	0,57
<i>SPAC14C4.15c</i>	1,52	<i>SPNCRNA.74</i>	0,57
<i>SPAC24H6.02c</i>	1,51	<i>SPAC630.04c</i>	0,58
<i>SPBC16A3.16</i>	1,51	<i>SPBPB21E7.10</i>	0,58
<i>SPAC6G10.10c</i>	1,51	<i>SPNCRNA.73</i>	0,6
<i>SPBC1289.17</i>	1,51	<i>SPAC56F8.12</i>	0,62
<i>SPBC1709.10c</i>	1,5	<i>SPBC359.02</i>	0,63
<i>SPAC19D5.11c</i>	1,49	<i>SPCC830.07c</i>	0,64
<i>SPAC56F8.10</i>	1,48	<i>SPNCRNA.69</i>	0,64
<i>SPNCRNA.21</i>	1,47	<i>SPAC227.15</i>	0,66
<i>SPAC227.17c</i>	1,47	<i>SPCC1753.02c</i>	0,67
<i>SPBC685.09</i>	1,46	<i>SPMIT.03</i>	0,67
<i>SPAC23C11.12</i>	1,45	<i>SPAC1B3.04c</i>	0,68
<i>SPCC162.11c</i>	1,45	<i>SPBC29A10.08</i>	0,68

<i>SPAC9E9.09c</i>	1,45	<i>SPBC359.01</i>	0,68
<i>SPBC14F5.12c</i>	1,44	<i>SPBC1D7.02c</i>	0,69
<i>SPBC3E7.06c</i>	1,44	<i>SPBC1861.02</i>	0,71
<i>SPBC15C4.02</i>	1,43	<i>SPBP19A11.02c</i>	0,71
<i>SPBC28F2.07</i>	1,42	<i>SPCC736.15</i>	0,72
<i>SPAC5H10.12c</i>	1,41	<i>SPAC14C4.11</i>	0,73
<i>SPCC11E10.06c</i>	1,41	<i>SPNCRNA.93</i>	0,74
<i>SPBC887.11</i>	1,41		
<i>SPAC1952.12c</i>	1,41		
<i>SPAC56F8.09</i>	1,4		
<i>SPBC1198.03c</i>	1,4		
<i>SPAC212.04c</i>	1,4		
<i>SPAC11H11.03c</i>	1,39		
<i>SPBC27B12.07</i>	1,39		
<i>SPBP23A10.04</i>	1,39		
<i>SPBC19F8.01c</i>	1,38		
<i>SPAC16A10.05c</i>	1,38		
<i>SPAC922.03</i>	1,37		
<i>SPAC23H4.16c</i>	1,36		
<i>SPCC417.11c</i>	1,36		
<i>SPAC17A5.01</i>	1,35		
<i>SPBC342.03</i>	1,35		
<i>SPAC24H6.01c</i>	1,35		
<i>SPAC13F5.01c</i>	1,35		
<i>SPCC11E10.08</i>	1,34		
<i>SPAC17A5.04c</i>	1,34		
<i>SPBC3B9.04</i>	1,34		
<i>SPAC343.07</i>	1,33		
<i>SPAC12G12.13c</i>	1,32		
<i>SPBP8B7.08c</i>	1,32		
<i>SPAC20G8.07c</i>	1,31		
<i>SPCC24B10.15</i>	1,3		
<i>SPBC887.12</i>	1,29		
<i>SPAC30D11.01c</i>	1,29		
<i>SPAC17H9.02</i>	1,25		

Supplementary Table S2. *S. pombe* strains used in this study.

Strain	Genotype	Reference
FBY101	h+ ade6-M210 leu1-32 ura4Δ18 his3Δ1 pab2-TAP::kanMX6	70
FBY106	h+ ade6-M216 leu1-32 ura4Δ18 his3Δ1	73
FBY107	h- ade6-M216 leu1-32 ura4Δ18 his3Δ1 pab2::kanMX6	73
FBY176	h+ leu1Δ ura4Δ his2Δ dis3-54	40
FBY185	h+ ade6-M210 leu1-32 ura4Δ18 his3Δ1 pab2-GFP::kanMX6	This study
FBY193	h+ leu1Δ ura4Δ his2Δ dis3-54 pab2::ura4	This study
FBY195	h- ade6-704 ura4Δ leu1-32 rrp6::kanMX6	69
FBY207	h+ ade6-M210 leu1-32 ura4Δ18 his3Δ1 rrp6-TAP::kanMX6	This study
FBY209	h+ ade6-M210 leu1-32 ura4Δ18 his3Δ1 dis3-TAP::kanMX6	This study
FBY233	h- ade6-704 ura4Δ leu1-32 rrp6::kanMX6 pab2::ura4	This study
FBY245	h- ade6-M210 leu1-32 ura4Δ18 his3Δ1 cid14::ura4	This study
FBY250	h- ade6-M216 leu1-32 ura4Δ18 his3Δ1 cid14::ura4 pab2::kanMX6	This study
FBY252	h+ ade6-M210 leu1-32 ura4Δ18 his3Δ1 ski7::ura4	This study
FBY253	h- ade6-M216 leu1-32 ura4Δ18 his3Δ1 ski7::ura4 pab2::kanMX6	This study
FBY288	h+ ade6-M210 leu1-32 ura4Δ18 his3Δ1 pab2-GFP::kanMX6 rrp6::ura4	This study
FBY341	h+ ade6-M210 leu1-32 ura4Δ18 his3Δ1 pab2::ura4 SPAC29A4.04c-TAP::kanMX6	This study

ARTICLE 2

THE NRD1-LIKE PROTEIN SEB1 COORDINATES CO-TRANSCRIPTIONAL 3' END PROCESSING AND POLYADENYLATION SITE DECISION

Jean-François Lemay¹, Samuel Marguerat², Marc Laroche¹, Xiaochuan Liu³, Rob van Nues⁴, Mainul Hoque³, Bin Tian³, Sander Granneman⁴, Jürg Bähler⁵ and François Bachand¹

En cours de révision à : Genes and Development

¹ RNA Group, Department of Biochemistry, Université de Sherbrooke, Sherbrooke, Quebec, Canada

² Medical Research Council Clinical Sciences Centre, Imperial College London, London, U.K.

³ Centre for Synthetic and Systems Biology and Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, U.K.

⁴ Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers New Jersey Medical School and Rutgers Cancer Institute of New Jersey, Newark, New Jersey, U.S.A.

⁵ Research Department of Genetics, Evolution and Environment, University College London and UCL Cancer Institute, London, U.K.

CONTRIBUTION

Cette étude est le résultat d'une collaboration multicentrique impliquant les laboratoires de recherche des professeurs Samuel Marguerat, Sander Granneman, Bin Tian, Jürg Bähler et François Bachand.

Ma contribution à cet article fut de (1) développer le projet en collaboration avec F.B., (2) créer la quasi-totalité des souches utilisées dans cette étude ainsi que (3) d'effectuer l'ensemble des expériences de ChIP, de buvardage de type *Northern*, de buvardage de type *Western* (incluant les expériences de co-immunoprécipitation) et d'essai de croissance. J'ai également écrit entièrement la version initiale du manuscrit. J'ai donc contribué de manière importante, tant d'un point de vue intellectuel que technique, à la complétion de ce projet. Sommaire des figures/expériences réalisées par J-F.L. : Fig. 1A-D; Fig. 2A-D; Fig. 3D, F, H; Fig. 4; Fig. 6B, D, F, H; Fig. S2; Fig. S4D-K; Fig. S5; Fig. S7A-K.

S.M fut responsable de l'analyse computationnelle des données dérivées des expériences de RNA-seq, de ChIP-seq et de CRAC. J.B. a supervisé la préparation des bibliothèques associées aux expériences de RNA-seq. R.v.N. a effectué les expériences de CRAC sous la supervision de S.G. X.L. a effectué les expériences de 3'READS et a analysé les données provenant de celles-ci. M.H. et X.L. ont préparé les bibliothèques associées aux expériences de 3' READS. B.T. a supervisé le travail de M.H et X.L. M.L. a effectué les expériences de ChIP suivantes : (1) Clp1-TAP et Cft2-TAP ainsi que (2) ARNPII (à des fins de ChIP-seq et de normalisation des résultats de ChIP de Clp1-TAP et Cft2-TAP). M.L. a également effectué le buvardage de type Northern associé à l'expérience où des sites de liaison putatifs de Seb1 sont mutés au niveau du gène rapporteur *snR99*, les expériences impliquant l'utilisation de 6-AU et a produit un nombre limité de souche.

RÉSUMÉ

La terminaison de la transcription est un processus intimement associé à la maturation de l'extrémité 3' d'un ARN. Pour les gènes codants pour des protéines, la terminaison est normalement initiée par un clivage endonucléolytique au site de polyadénylation par la machinerie de clivage et de polyadénylation. Contrastant avec ce mécanisme, la terminaison transcriptionnelle de gènes non-codants chez la levure à bourgeon *S. cerevisiae* ne requiert pas la machinerie de maturation 3' des ARNms, mais plutôt le complexe NNS (Nrd1-Nab3-Sen1). Ici, nous présentons des données arguant contre la conservation de ce mécanisme de terminaison chez l'organisme modèle *S. pombe*. Nos résultats indiquent plutôt que l'orthologue présumé de Nrd1 chez *S. pombe*, Seb1, est intimement lié au mécanisme de terminaison transcriptionnelle des gènes codants en dictant la sélection de site de polyadénylation sur l'ensemble du génome. Nos données révèlent que Seb1 s'associe à plusieurs composantes de la machinerie de clivage et de polyadénylation, est enrichie préférentiellement à l'extrémité 3' des gènes et lie des motifs d'ARN présents en aval des sites de clivage. Notamment, la déplétion de Seb1 engendre un accroissement généralisé de la longueur des 3' UTRs une conséquence de l'utilisation accrue de sites de polyadénylation alternatifs. Compte tenu que Seb1 affecte également le recrutement de facteurs liés à la maturation 3' des ARNs, nos résultats suggèrent que Seb1 promeut la réalisation d'évènements de maturation co-transcriptionnels impliqués dans la sélection de site de polyadénylation.

ABSTRACT

Termination of RNAPII transcription is associated with RNA 3' end formation. For coding genes, termination is initiated by the cleavage/polyadenylation machinery. In contrast, a majority of noncoding transcription events in *S. cerevisiae* do not rely on RNA cleavage for termination, but instead terminate via a pathway that requires the Nrd1-Nab3-Sen1 (NNS) complex. Here we show that the *S. pombe* ortholog of Nrd1, Seb1, does not function in NNS-like termination, but promotes polyadenylation site selection of coding and noncoding genes. We found that Seb1 associates with 3' end processing factors, is enriched at the 3' end of genes, and binds RNA motifs downstream of cleavage sites. Importantly, a deficiency in Seb1 resulted in widespread changes in 3' UTR length as a consequence of increased alternative polyadenylation. Given that Seb1 levels affected the recruitment of conserved 3' end processing factors, our findings indicate that the conserved RNA-binding protein Seb1 co-transcriptionally controls alternative polyadenylation.

INTRODUCTION

Termination of RNA polymerase II (RNAPII) transcription is a critical step of gene expression that is functionally associated with RNA 3' end formation and release of the nascent transcript from the site of transcription. For most protein-coding genes, current data suggest a model where transcription termination is initiated by the co-transcriptional recruitment of 3' end processing factors to the carboxy-terminal domain (CTD) of the largest subunit of RNAPII (Porrua et Libri, 2015). The CTD consists of a succession of heptad repeats, with the consensus amino acid sequence Y-S-P-T-S-P-S, which is subjected to a plethora of stage-dependent post-translational modifications that control the recruitment of various RNA processing factors (Cox et Mann, 2008). One of these CTD modifications, phosphorylation of serine 2 (Ser2P), gradually increases as transcription elongation progresses, and peaks at the 3' end of mRNA-encoding genes. Ser2P is in fact important for the recruitment of 3' end processing factors in both budding yeast (*Saccharomyces cerevisiae*) and human cells (Schonemann *et al.*, 2014), consistent with the overall conservation of 3' end processing factors (Xiang *et al.*, 2014). The transfer of 3' end processing factors from the transcription complex onto the nascent transcript involves the recognition of a functional poly(A) signal (PAS), composed of *cis*-acting RNA elements that define the site of pre-mRNA cleavage (Shi et Manley, 2015). The nascent pre-mRNA is subsequently cleaved by an endonuclease (Ysh1 in *S. cerevisiae*; CPSF73 in humans), generating a free 3' end for the polyadenylation machinery. Endonucleolytic cleavage also generates an uncapped 5' end to the RNA downstream of the cleavage site, providing an entry point for a 5'-3' exonuclease (Rat1 in *S. cerevisiae*; Xrn2 in humans) that has been proposed to chase RNAPII and promote its dissociation from the DNA template, a mechanism of transcription termination referred to as the torpedo model (Porrua et Libri, 2015).

In addition to mRNA-coding genes, RNAPII is also responsible for the synthesis of many noncoding RNAs (ncRNA), including small nuclear RNA (snRNAs), small nucleolar RNAs (snoRNAs), and cryptic unstable transcripts (CUTs). In *S. cerevisiae*, transcription termination of these ncRNAs does not rely on the 3' end processing machinery, but on a

mechanism that requires the activity of the Nrd1-Nab3-Sen1 (NNS) complex (Porrúa et Libri, 2015). Furthermore, whereas termination at protein-coding genes correlates with Ser2 phosphorylation, NNS recruitment to the CTD is primarily driven by Ser5 phosphorylation via the CTD-interaction domain of Nrd1 (Cox et Mann, 2008, Porrúa et Libri, 2015). The presence of RNA sequence motifs is crucial to subsequently engage the NNS complex onto nascent ncRNAs via the RNA-binding properties of Nrd1 and Nab3 (Creamer *et al.*, 2011, Schulz *et al.*, 2013, Webb *et al.*, 2014). Interestingly, transcription termination by the NNS complex does not appear to be associated with endonucleolytic cleavage of the nascent RNA, but rather functions by a mechanism that dislodges RNAPII from the DNA template via the helicase activity of Sen1 (Porrúa et Libri, 2013). Another important feature of NNS-dependent termination is its functional association with the exosome complex of 3'-5' exonucleases, which contributes to 3' end trimming of snRNA/snoRNA precursors and to the rapid degradation of products of pervasive transcription (Arigo *et al.*, 2006, Thiebaut *et al.*, 2006, Vasiljeva et Buratowski, 2006).

Interestingly, genome-wide studies indicate that pervasive transcription is widespread not only in *S. cerevisiae*, but also in many other species, including humans (Chan *et al.*, 2014). Yet, despite the many studies that have underscored the critical role of the *S. cerevisiae* NNS complex in limiting the extent of pervasive transcription, the conservation of NNS-like transcription termination across eukaryotic species has remained elusive. Human SCAF8 and *Schizosaccharomyces pombe* Seb1 share common protein domain architecture with *S. cerevisiae* Nrd1, including a conserved RNA recognition motif and a CTD-interacting domain (Meinhart et Cramer, 2004). Fission yeast Seb1 was in fact shown to function in heterochromatin assembly via binding to ncRNAs originating from pericentromeric repeats (Marina *et al.*, 2013), which is reminiscent of the silencing function of Nrd1 at ribosomal DNA and telomeric loci (Vasiljeva *et al.*, 2008). As yet, however, it remains unknown whether *S. pombe* Seb1 and human SCAF8 function in NNS-like transcription termination.

Here, we set out to characterize NNS-like transcription termination in *S. pombe* and disclose its functional relevance in transcriptome surveillance. Unexpectedly, proteomic

analysis of Seb1-associated proteins found no evidence for an NNS-like complex, but rather identified several mRNA 3' end processing factors. Furthermore, transcriptome-wide analysis of Seb1-RNA associations revealed widespread binding downstream of polyadenylation sites. The functional significance of this RNA binding pattern was demonstrated by increased levels of alternative polyadenylation in Seb1-depleted cells, resulting in global changes in 3' UTR lengths. Our data suggest that Seb1 controls poly(A) site selection by promoting the recruitment of specific cleavage/polyadenylation factors at the 3' end of genes via a mechanism linked to transcription elongation kinetics. Our findings reveal that regulation of 3' UTR length is a co-transcriptional process controlled by the recruitment of the Seb1 RNA-binding protein at the 3' end of genes.

RESULTS

Global RNAPII transcription termination defects in Seb1-depleted cells

To examine whether an NNS-like complex exist in *S. pombe*, amino-acid sequence comparisons against the fission yeast proteome identified gene products with substantial sequence homology to *S. cerevisiae* Nab3 and Nrd1: 22% and 29% identical (53% and 57% similar) to *S. pombe* SPAC3H8.09c (Nab3) and SPAC222.09 (Seb1), respectively. Interestingly, the fission yeast genome expresses two distinct Sen1 paralogs, SPAC6G9.10c (Sen1) and SPBC29A10.10c (Db18), which show 26% and 27% identity (58% similarity), respectively, to *S. cerevisiae* Sen1. Surprisingly, only the *NRD1* homolog, *seb1*, is essential for viability in *S. pombe*, which contrasts to *S. cerevisiae*, where *NRD1*, *NAB3*, and *SEN1* are all essential genes. We also found that the *S. pombe* *dbl8Δ/sen1Δ* double mutant was viable (data not shown).

The *S. cerevisiae* NNS complex is well known for its involvement in termination of noncoding transcripts such as snoRNA genes (Porrua et Libri, 2015). To test if this function is conserved in fission yeast, we measured RNAPII density along a snoRNA gene (Fig. 1a) by chromatin immunoprecipitation (ChIP) assay in strains in which genes encoding for

putative orthologs of the budding yeast NNS complex were either deleted (*nab3*, *sen1*, or *dbl8*) or expressed under the control of the thiamine-sensitive *nmt1* promoter in the case of the essential *seb1* gene (P_{nmt1} -*seb1*). In wild-type cells, RNAPII ChIP signals showed the expected gradual decline as transcription progresses downstream of the *snR3* gene (Fig. 1b). Deletion of *nab3*, *sen1*, *dbl8* as well as the *dbl8Δ/sen1Δ* double deletion did not markedly affect this RNAPII profile (Fig. 1b). In contrast, the depletion of Seb1 in thiamine-supplemented medium resulted in increased levels of RNAPII at the 3' end of *snR3* relative to the wild-type strain grown in the same conditions (Fig. 1b, see regions 4-5), consistent with transcription termination defects. A similar RNAPII profile was observed after Seb1 depletion using a CTD-independent ChIP approach that used a strain expressing an HA-tagged version of a core RNAPII component, Rpb3 (Fig. 1c). Read-through transcription was also observed by analyzing the distribution of elongating RNAPII by a transcription run-on (TRO) assay, revealing increased production of nascent RNA at the 3' end of *snR3* in Seb1-depleted cells (Fig. 1d,e). Therefore, a substantial proportion of RNAPII that terminates downstream of *snR3* in wild-type cells fails to terminate in Seb1-deficient cells. Importantly, global analysis of RNAPII levels in Seb1-depleted cells by ChIP-seq revealed widespread transcription termination defects that were not limited to noncoding genes (Fig. 1f), but also apparent at protein-coding genes (Fig. 1g,h). Plotting the cumulative levels of RNAPII binding relative to annotated cleavage/polyadenylation sites indicated a clear shift in polymerase density downstream of the noticeable decline observed in wild-type cells (Fig. 1i). This demonstrates that transcription termination defects are a general feature of Seb1-depleted cells. Importantly, this cumulative RNAPII profile of Seb1-depleted cells was not biased by a specific class of gene nor restricted to a particular genomic arrangement, as mRNA-, snoRNA- and snRNA-encoding genes all showed transcription termination defects as well as both tandem and convergent gene pairs (Fig. S1). These results indicate that a deficiency in Seb1 causes widespread defects in RNAPII termination at coding and noncoding genes, which contrasts to *S. cerevisiae* Nrd1 that mainly functions at noncoding RNAs. The functional divergence between *S. pombe* Seb1 and *S. cerevisiae* Nrd1 was further demonstrated by the inability of Seb1 to complement loss of Nrd1 functions in *S. cerevisiae*, while expression of Nrd1 did not complement loss of Seb1 functions in *S. pombe* (Fig. S2). We conclude that Seb1 and Nrd1 are not functional homologs.

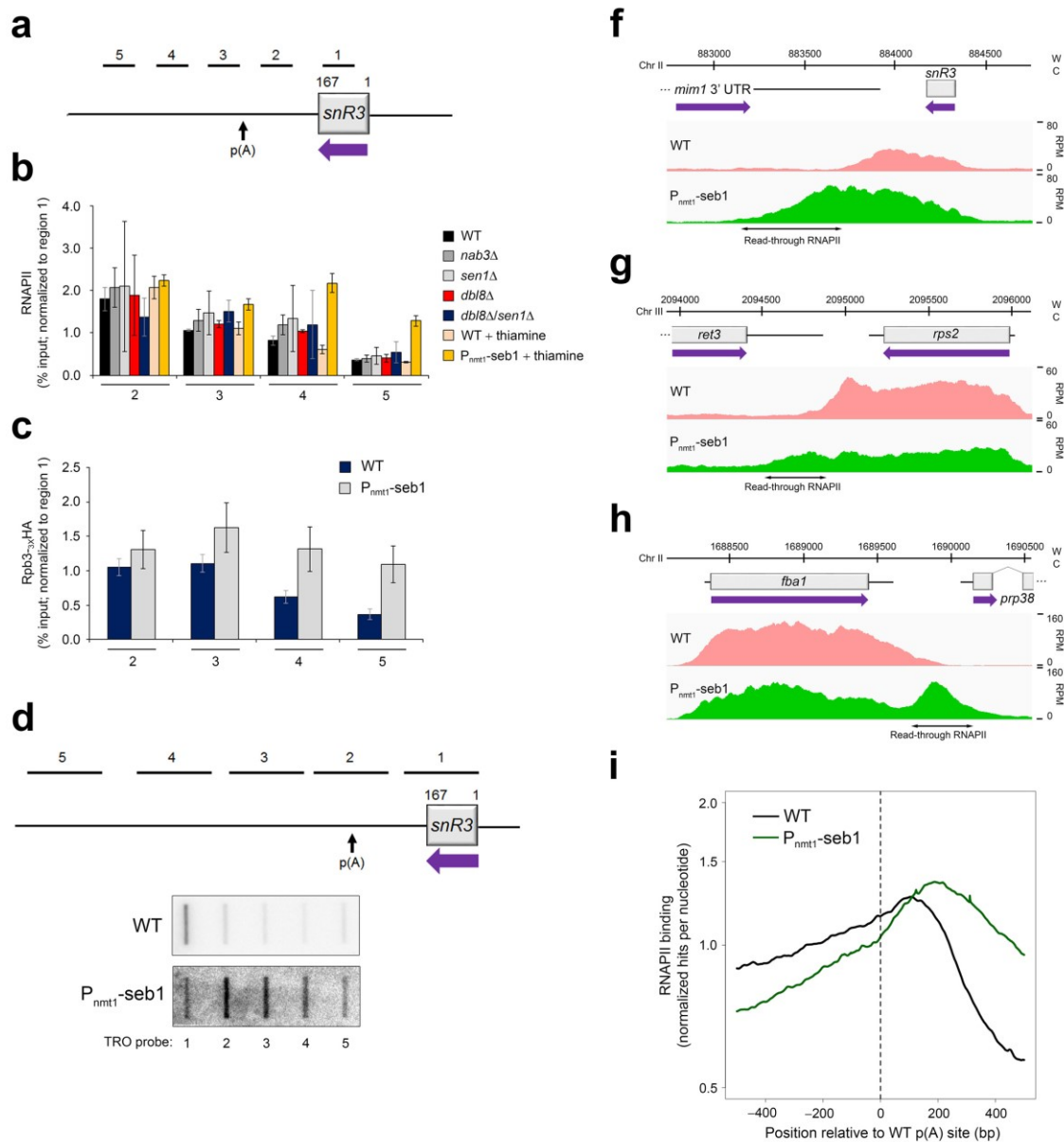


Figure 1. Transcription termination defects in Seb1-depleted cells.

(a). Schematic of the *snR3* snoRNA locus. Bars above the gene show the positions of PCR products used for ChIP analyses in panels b-c. p(A) refers to the polyadenylation site of the 3'-extended precursor (Lemay *et al.*, 2010). (b) ChIP analyses of RNAPII density along the *snR3* gene using extracts prepared from wild-type (WT) and the indicated mutant strains. ChIP signals (% Input) were normalized to region 1. Error bars, s.d. (n=3 biological replicates from independent cell cultures). (c) ChIP analyses of HA-tagged Rpb3 (Rpb3-_{3x}HA) along the *snR3* gene in Seb1-depleted cells (*P_{nmt1}-seb1*) or in control (WT) cells. Error bars, s.d. (n=3 biological replicates from independent cell cultures). (d) Schematic showing the position of probes (1-5) used for transcription run-on (TRO) assays along the *snR3* snoRNA locus. (e) Representative TRO blot for *snR3*. (f, g, h) RNAPII profiles (ChIP-seq) across the *snR3* (f), *rps2* (g) and *fba1* (h) genes for the indicated strains. W: Watson strand; C: Crick

strand; RPM: Reads Per Million. (i) Cumulative RNAPII profile relative to poly(A) sites in the indicated strains. Curves show the sum of normalized ChIP-seq sequencing scores over a genomic region covering the major poly(A) site.

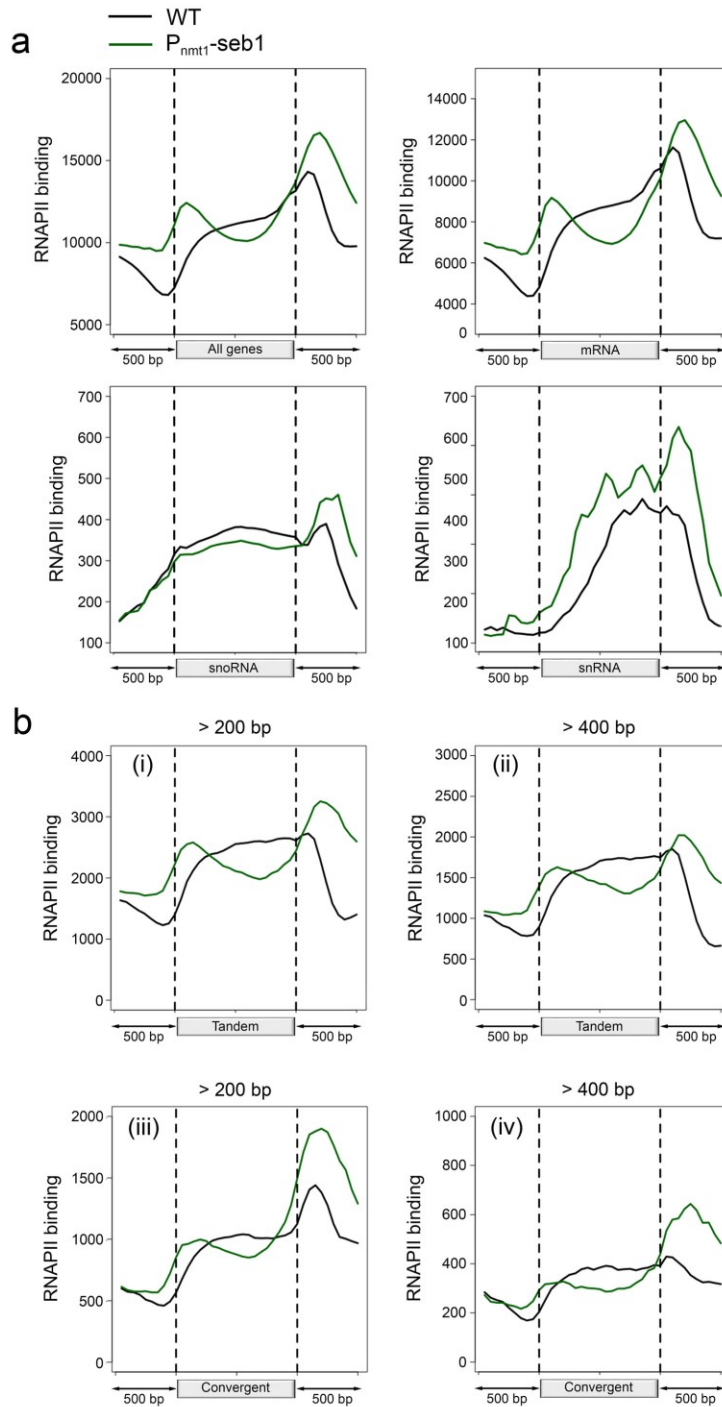


Figure S1. Read-through transcription by RNAPII in Seb1-deficient cells.

(a) Average distribution of RNAPII density in wild-type (WT) and Seb1-depleted cells (P_{nmt1} -seb1), as determined by analysis of ChIP-seq results for all ($n = 7005$), mRNA ($n = 5123$), snoRNA ($n = 53$) and snRNA ($n = 7$) genes. Light grey rectangles represent open reading frames (mRNA genes) or noncoding sequences (snoRNA and snRNA genes) and black arrows 500 base pairs (bp) of upstream and downstream flanking DNA sequences. (b) As in (a) but for (i) tandem genes that are more than 200 bp apart ($n = 1425$), (ii) tandem genes that are more than 400 bp apart ($n = 855$), (iii) convergent genes that are more than 200 bp apart ($n = 452$) as well as (iv) convergent genes that are more than 400 bp apart ($n = 208$).

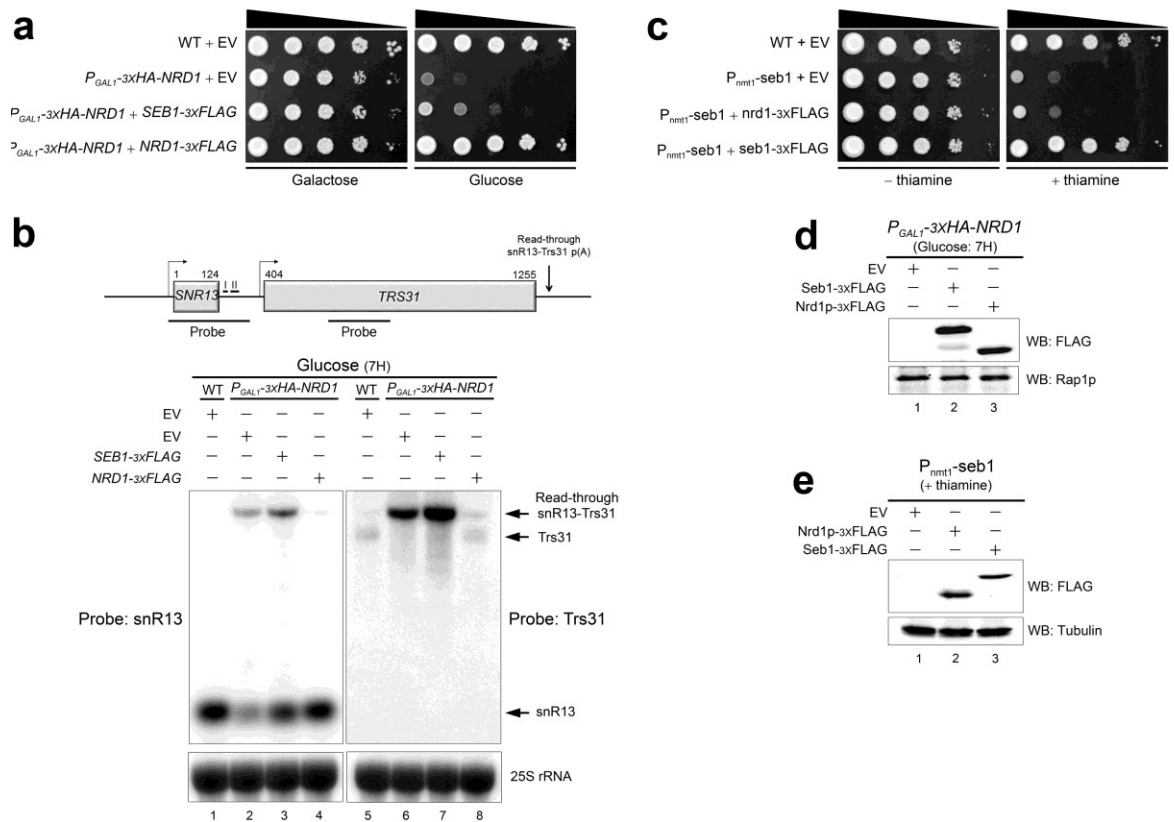


Figure S2. Seb1 is not the functional ortholog of Nrd1.

(a) Ten-fold serial dilutions of *S. cerevisiae* wild-type (WT) cells transformed with an empty vector (EV) as well as P_{GALI} -3XHA-NRD1 cells that were transformed with an EV or constructs that express a 3XFLAG-tag wild-type version of Seb1 or a 3XFLAG-tag wild-type version of Nrd1p. Cells were spotted on galactose-containing and glucose-containing media to respectively allow and repress transcription from the P_{GALI} promoter. (b) Upper part: Schematic representation of the *SNR13-TRS31* genomic environment. Bars under the genes show the positions of strand-specific RNA probes used for northern blot analyses. The position of the Trs31/snR31-Trs31 p(A) is illustrated as well as the regions (I/II) known to harbor *cis*-elements directing termination of the *SNR13* gene (Steinmetz et Brow. 2003).

Lower part: northern blot analyses of snR13 (lanes 1-4) and Trs31 (lanes 5-8) prepared from wild-type (WT) cells transformed with an empty vector (EV) (lanes 1 and 5) or from *P_{GALI}-3XHA-NRDI* cells that were transformed with an EV (lanes 2 and 6) or constructs that express a 3XFLAG-tag wild-type version of Seb1 (lanes 3 and 7) or a 3XFLAG-tag wild-type version of Nrd1p (lanes 4 and 8). The 25S rRNA serves as a loading control. (c) Ten-fold serial dilutions of wild-type (WT) cells transformed with an empty vector (EV) as well as *P_{nmt1}-seb1* cells that were transformed with an EV or constructs that express a 3XFLAG-tag wild-type version of Seb1 or a 3XFLAG-tag wild-type version of Nrd1p. Cells were spotted on minimal medium lacking or containing thiamine to respectively allow and repress transcription from the *P_{nmt1}* promoter. (d) Immunoblot analysis of whole-cell extracts prepared from *P_{GALI}-3XHA-NRDI* cells that were transformed with an EV (lane 1) or constructs that express a 3XFLAG-tag wild-type version of Seb1 (lane 2) or a 3XFLAG-tag wild-type version of Nrd1p (lane 3). Rap1p protein serves as a loading control. (e) As in (d) but in *P_{nmt1}-seb1* background cells. Tubulin serves as a loading control. Cells were either grown in glucose (b,d) or thiamine-containing medium (e).

Seb1 associates with proteins involved in mRNA 3' end processing and is enriched at the 3' end of genes

To elucidate the mechanism by which a deficiency in Seb1 results in transcription termination defects, we decided to identify Seb1-associated proteins using a functional His-TEV-Protein A-tagged version of Seb1 (Seb1-HTP) expressed from its endogenous chromosomal locus. Purification of Seb1 (Fig. 2a) coupled to mass spectrometry resulted in the identification of 471 Seb1-associated proteins (Table S1). Consistent with the conclusion that Seb1 does not function in NNS-like transcription termination, Nab3- and Sen1-specific peptides were not identified in the Seb1 purification, while only 3 Dbl8-specific peptides were detected (2.1% sequence coverage). We next used computer algorithms (Berriz *et al.*, 2009) to distinguish functional protein classes within the top 10% of the Seb1-associated proteins. Notably, a significant number of proteins involved in mRNA 3' end processing were enriched among the top 10% ($p = 7.96^{-6}$, Fisher's Exact Test; Fig. 2b; Table S2): Rna14, from the cleavage factor 1A (CF1A) complex, as well as Cft1, Cft2, Ysh1, and Pta1, from the cleavage and polyadenylation factors (CPF) complex (Xiang *et al.*, 2014). Reciprocal immunoprecipitation assays confirmed the nuclease-resistant association between Seb1 and Ysh1 (Fig. 2c) as well as between Seb1 and Cft2 (Fig. 2d). Collectively, 18 components

predicted to be part of the *S. pombe* cleavage and polyadenylation machinery were copurified with Seb1 (Table S2).

The purification of Seb1 also identified factors predicted to function in transcription termination, such as Dhp1, Rhn1, and Din1 (Fig. 2b and Table S2), which share homology to Rat1, Rtt103, and Rai1, respectively, three proteins that are required for termination of RNAPII transcription in *S. cerevisiae* (Kim *et al.*, 2004). In addition, core components of RNA polymerase I and II complexes (Fig. 2b; $p = 2.552^{-5}$, Fisher's Exact Test) as well as a significant number of proteins involved in ncRNA processing (Fig. 2b; $p = 3.812^{-13}$, Fisher's Exact Test) were identified in the top 10% of Seb1-associated proteins.

Binding profiles of mRNA 3' end processing factors by ChIP tend to show recruitment at the 3' end of genes (Schonemann *et al.*, 2014). Consistent with the copurification of Seb1 with mRNA 3' end processing factors, ChIP-seq analysis of Seb1 indicated strong enrichment at the 3' end of snoRNA- (Fig. 2e) and mRNA-encoding genes (Fig. 2f-g), with genome-wide cross-linking in the vicinity of poly(A) sites (Fig. 2h). Notably, the 3' end enrichment profile of Seb1 was irrespective of gene length or gene classes (Fig. 2h; Fig. S3a-d), consistent with the generalized transcription termination defects observed in Seb1-deficient cells (Fig. 1). In contrast, a 3' end enrichment of Seb1 was not observed at RNAPI- and RNAPIII-transcribed genes (Fig. S3e-f), suggesting a mode of 3' end recruitment specific to RNAPII transcription. Together, the protein interaction network and the ChIP-seq analysis of Seb1 strongly support a general role in 3' end processing and transcription termination.

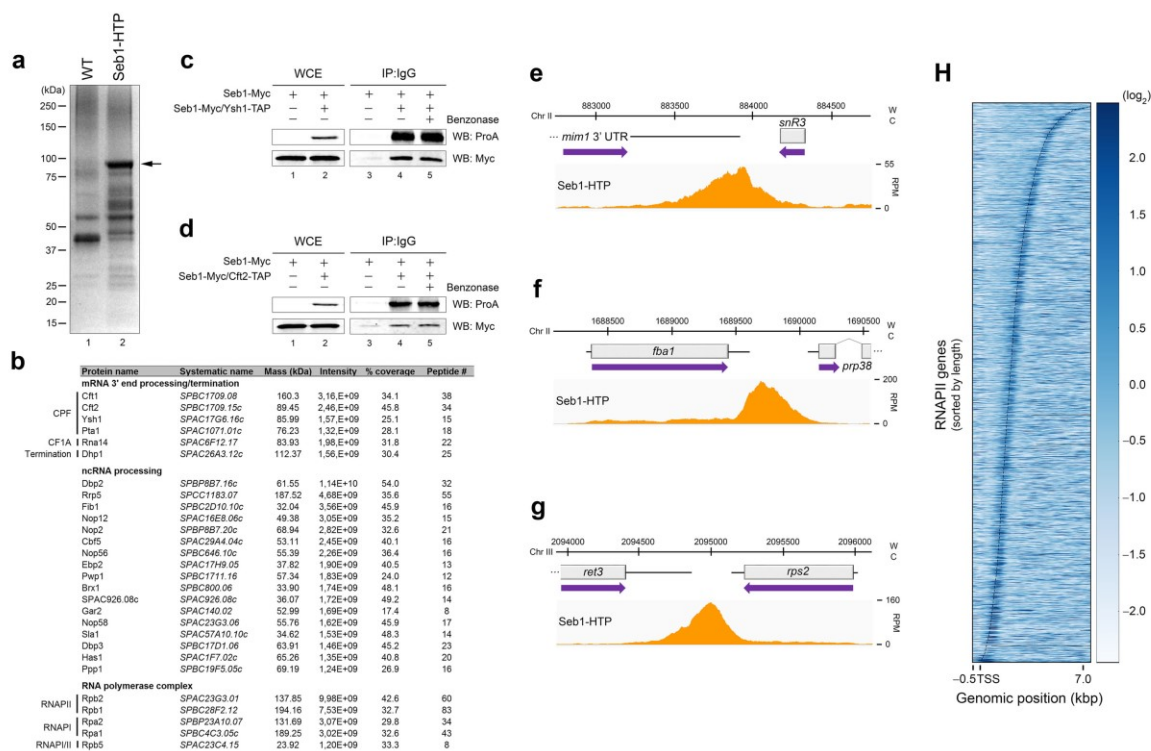


Figure 2. Seb1 interacts with the 3' end processing machinery and is enriched at the 3' end of genes.

(a) Coomassie blue staining of proteins co-purified with Seb1-HTP (lane 2) and from a control untagged strain (lane 1). The arrowhead indicates the position of Seb1-HTP. (b) A subset of the top 10% of Seb1-associated proteins identified by LC-MS/MS is shown. The intensity represents the relative abundance (peptide intensity), while the % coverage and the peptide # represent the unique peptide sequence coverage and the number of unique peptides, respectively. (c,d) Immunoblot analyses of whole-cell extracts (WCE; lanes 1-2) and IgG-sepharose precipitates (IP: IgG; lanes 3-5) prepared from control Seb1-Myc cells or Seb1-Myc cells co-expressing a TAP-tag version of Ysh1 (c) or Cft2 (d). Purification experiments were performed in the absence or presence of the Benzonase nuclease (lanes 4-5). (e-g) ChIP-seq analysis of Seb1-HTP occupancy along the *snR3* (e), *fba1* (f) and *rps2* (g) genes. W: Watson strand; C: Crick strand; RPM: Reads Per Million. (h) Heat-map of Seb1 DNA-binding sites derived from ChIP-seq for all RNAPII-transcribed genes. Genes were sorted by length and aligned at their transcription start site (TSS). The curved line represents the p(A) sites. Strength of binding is coded from white (no binding) to dark blue (strong binding).

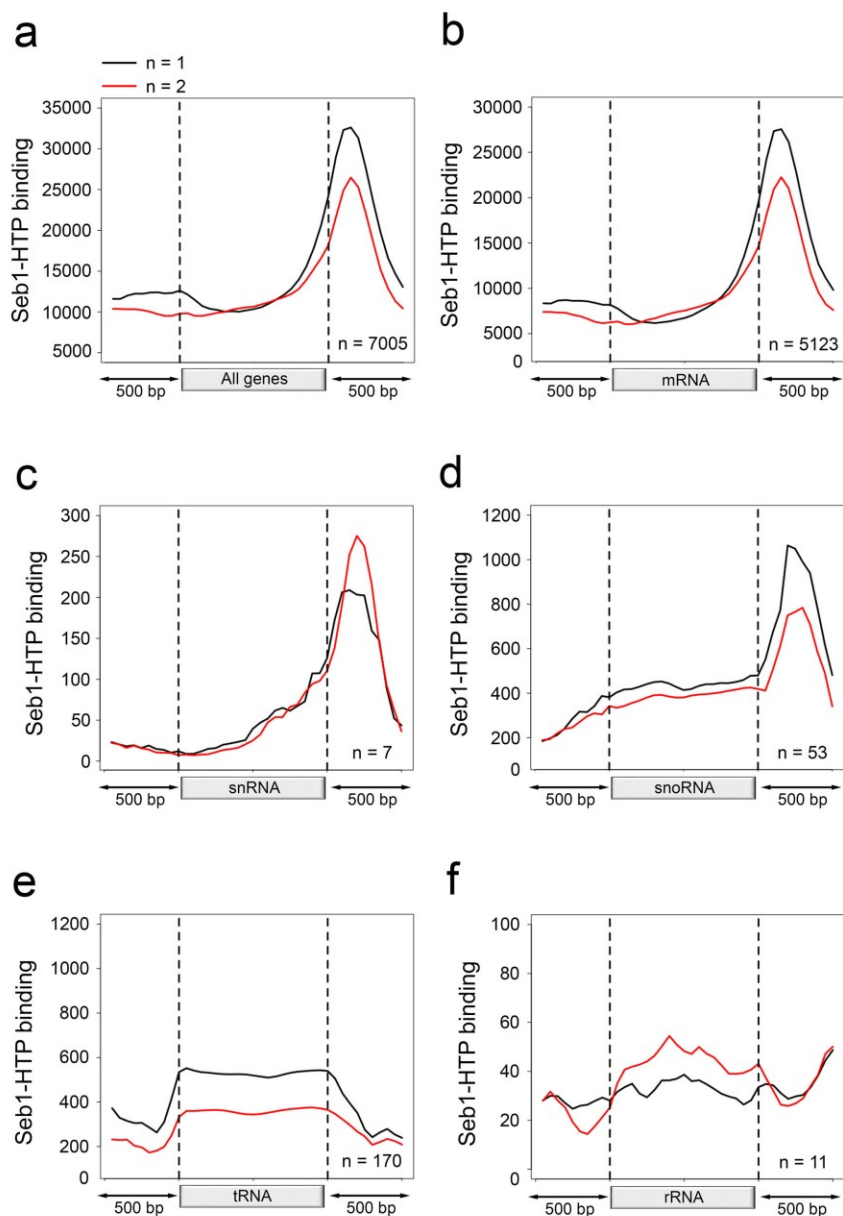


Figure S3. Genome-wide distribution of Seb1 at the 3' end of RNAPII genes.

(a-f) Average gene distribution of Seb1-HTP density as determined by analysis of results from two independent ChIP-seq experiments for (a) all, (b) mRNA, (c) snRNA, (d) snoRNA, (e) tRNA and (f) rRNA genes. Light grey rectangles represent open reading frames (mRNA genes) or noncoding sequences (snoRNA, snRNA, tRNA and rRNA genes) and black arrows 500 base pairs (bp) of upstream and downstream flanking DNA sequences. Sample size for each gene classes is indicated at the bottom right of the graphs. All the experiments in Figure S3 were performed from cells grown in YES medium.

Seb1 controls poly(A) site selection

According to current models of transcription termination (Porrua et Libri, 2015), at least two mechanisms could account for the generalized increased in read-through transcription in *Seb1*-deficient cells. First, *Seb1* could promote 3' end processing (PAS recognition and/or mRNA cleavage), which according to the torpedo model, would result in termination defects in conditions of *Seb1* deficiency. Alternatively, *Seb1* could function subsequent to mRNA cleavage by promoting dissociation of RNAPII from the DNA template following passage through PAS elements. If *Seb1* functions after pre-mRNA cleavage, 3' end processing defects are not expected; in contrast, if *Seb1* regulates pre-mRNA cleavage, defects in 3' end processing are expected in *Seb1*-depleted cells. To distinguish between these models, we compared the landscape of poly(A) site selection between *Seb1*-deficient and control cells by 3' Region Extraction And Deep Sequencing (3' READS), an approach developed to map mRNA cleavage sites at the genome-wide level (Hoque *et al.*, 2013). Out of 5393 genes with mappable p(A) sites, we found that 3013 genes (55%) showed a change in p(A) site decision in *Seb1*-depleted cells relative to the control (Fig. 3a): 2484 genes (82%) showed lengthening of the 3'-most exon (red) (Table S3), whereas 529 genes (18%) had shorter 3'-most exons (blue). The preferential use of distal p(A) sites in the *seb1* mutant resulted in a significant increase in median 3' UTR length (Fig. 3b): from 168-nt in wild-type to 256-nt in *Seb1*-deficient cells. Importantly, the *seb1* mutant showed increased heterogeneity in p(A) site positions relative to wild-type cells: analysis of relative abundance for the top-three p(A) sites (based on abundance in wild-type cells) for each gene indicated a significant decrease for the most abundant p(A) site in the *seb1* mutant relative to the wild-type (Fig. S4a); in contrast, secondary p(A) sites were more frequently used in the *seb1* mutant (Fig. S4b-c). Examples of coding and noncoding genes whose cleavage site positions were strongly affected by *Seb1* deficiency are shown in Fig. 3c-h and in Fig. S4d-k. In eukaryotes, alternative polyadenylation (APA) has the potential to influence the stability, translation, and localization of a given mRNA through the inclusion or exclusion of *cis*-acting elements in the 3' UTR (Tian et Manley, 2013). We thus examined whether the changes in p(A) site position observed in the *seb1* mutant influenced mRNA abundance by analyzing standard RNA-seq data. Interestingly, the 3' UTR lengthening phenotype detected in the *seb1*

mutant did not markedly change mRNA abundance (Fig. 3i). In sum, these results indicate that Seb1 functions in poly(A) site selection.

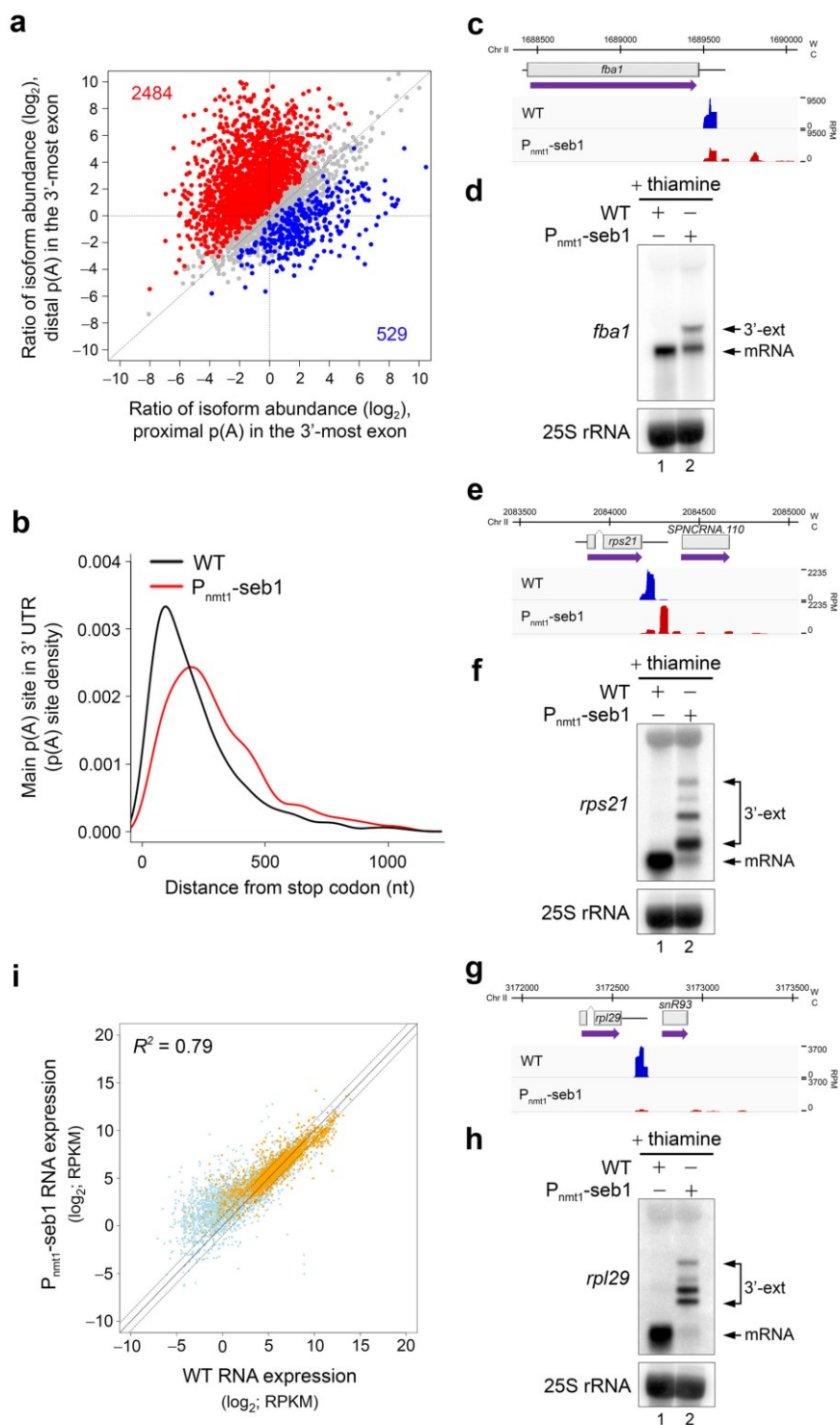


Figure 3. *Seb1* levels affect poly(A) site selection.

(a) Regulation of alternative p(A) site utilization (APA) in the 3'-most exon as determined by 3'READS. The number of genes with significantly lengthened 3' UTR (red dots) and the number of genes with significantly shortened 3' UTR (blue dots) is indicated in the graph. Significantly regulated isoforms are those with p -value < 0.05 (Fisher's exact test). Only the two most abundant isoforms for each gene were analyzed. (b) Distribution of 3'READS-derived p(A) sites relative to the upstream stop codon. Major p(A) site is the one with the highest number of reads per gene. Mapped p(A) sites in *Seb1*-depleted cells, as compared to wild-type cells (WT), are on average significantly more distant from the stop codon (p -value $< 2.22 \cdot 10^{-16}$ by Wilcoxon rank-sum test). (c-h) 3'READS profiles and northern blot analyses of *fbal* (c-d), *rps21* (e-f), and *rpl29* (g-h) genes in the indicated strains. W: Watson strand; C: Crick strand; RPM: Reads Per Million. 3'-extended transcripts that accumulate in *Seb1*-depleted condition are shown (3'-ext). (i) RNA expression changes for transcripts with 3'-extended 3' UTRs in WT and *Seb1*-depleted cells as measured by RNA-seq. RPKM, Reads Per Kilobase Per Million. The coefficients of determination (R^2) is indicated.

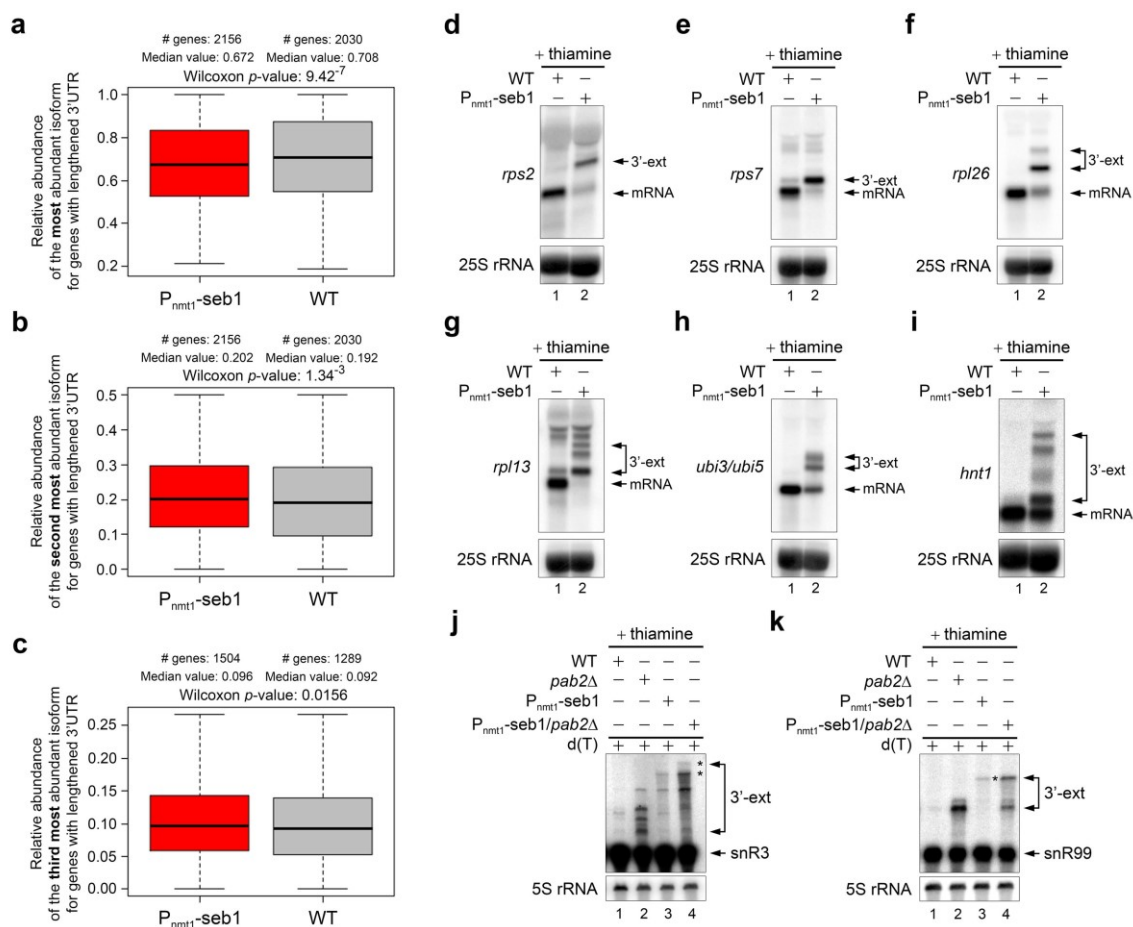


Figure S4. Alternative p(A) sites (APA) utilization in Seb1-deficient cells.

(a-c) Isoform relative abundance (RA) for the top 3 p(A) site ((a): most abundant; (b): second most abundant; (c): third most abundant) for each gene in wild-type (WT) and Seb1-depleted cells ($P_{nmt1-seb1}$). The RA of each isoform is shown as a fraction of the combined abundance for the top 3 p(A) site. The number of genes used to determine the RA of each isoform, the median RA value in each strain and the degree of significance between the WT and $P_{nmt1-seb1}$ results are all indicated at the top of each panel. Basically, the data show that there is a decrease of RA for the most abundant p(A) site in the *seb1* mutant compared to the WT and, consequently, increases of RA for the other 2 p(A) sites in the mutant. (d-i) Northern blot analyses of the *rps2* (d), *rps7* (e) *rpl26* (f), *rpl13* (g), *ubi3/ubi5* (h) and *hnt1* (i) mRNA-encoding genes using RNA from WT and $P_{nmt1-seb1}$ cells. 3'-ext refers to the 3'-extended transcripts that accumulate in Seb1-depleted condition. The 25S rRNA is used as a loading control. (j-k) Northern blot analyses of snoRNA-encoding genes using RNA obtained from the indicated strains. Total RNA was treated with RNase H in the presence of a DNA oligonucleotide complementary to H/ACA class snoRNA snR99 (j) or snR3 (k). RNase H reactions were performed in the presence of oligo(dT) to collapse the 3'-extended polyadenylated forms (3'-ext) of these snoRNAs which are normally produced as part of their maturation/degradation cycle (Larochelle et al. 2012; Lemay et al. 2010). The *pab2Δ* and $P_{nmt1-seb1/pab2Δ}$ strains were used to increase the abundance of these 3'-ext forms (Lemay et al. 2010). The stars (*) indicate the 3'-ext forms that appear specifically following the depletion of Seb1. The 5S rRNA serves as a loading control. All experiments presented in Figure S4 are from cells grown in thiamine-containing medium.

The CID and RRM domains of Seb1 are required for poly(A) site selection

Although our results suggest that Seb1 is not the functional homolog of *S. cerevisiae* Nrd1, Seb1 shares extensive homology with the CTD-interaction domain (CID; 34% identity/70% similarity) and the RNA recognition motif (RRM; 46% identity/76% similarity) of Nrd1 (Fig. S5a). To test whether the predicted CID domain of Seb1 is necessary for normal 3' end processing/transcription termination, we took advantage of previously determined CTD-CID structures (Kubicek *et al.*, 2012, Lunde *et al.*, 2010, Meinhart et Cramer, 2004). In these structures, the CTD adopts a β -turn conformation that docks into a CID hydrophobic pocket via a set of conserved residues which, based on sequence alignment, would involve tyrosine (Y) 64, aspartic acid (D) 67, and arginine (R) 71 of Seb1 (Fig. S5a). We therefore generated a mutant allele that expressed a version of Seb1 (CID_{mut}) with substitutions at these particular residues. Wild-type and CID_{mut} alleles of *seb1* were chromosomally integrated as a single copy into the $P_{nmt1-seb1}$ conditional strain, and the extent to which the CID mutant restored the transcription termination defects induced by depletion of endogenous Seb1 was examined by RNAPII ChIP

assays along the *rps2* model gene (Fig. 4a). It should be noted that the converging *ret3* gene is barely transcribed by RNAPII (Fig. 1g), and is therefore unlikely to contribute to the ChIP signal downstream of *rps2*. As can be seen in Fig. 4b, the CID_{mut} version of Seb1 showed increases in RNAPII levels at the 3' end of *rps2* that were similar to Seb1-depleted cells (P_{nmt1-seb1} + EV), suggesting that the ability of Seb1 to interact with the CTD of Rpb1 is important for its function in 3' end processing. This conclusion was supported by RNA analyses, demonstrating altered p(A) site selection in cells that expressed the CID_{mut} version of Seb1 (Fig. 4c, lane 4).

In budding yeast, the preferential binding site of Nrd1 is formed after Ser5 phosphorylation of the CTD (Gudipati *et al.*, 2008, Vasiljeva *et al.*, 2008). To test the possibility that Ser5 phosphorylation is important for Seb1 recruitment and function, we expressed a version of Seb1 with amino acid substitutions at conserved residues (Ser5_{mut}; Fig. S5a) that were previously shown for *S. cerevisiae* Nrd1 to significantly decrease binding to Ser5-P CTD and that were necessary for Nrd1-dependent RNA processing (Kubicek *et al.*, 2012). Strikingly, the Ser5_{mut} version of Seb1 fully restored the defects in transcription termination (Fig. 4b) and in 3' end processing (Fig. 4c, lane 6) induced by the depletion of endogenous Seb1, arguing against a role for Ser5 phosphorylation in Seb1 recruitment and function.

Because Seb1 was previously shown to associate with ncRNA (Marina *et al.*, 2013), we examined whether its function in 3' end processing required RNA recognition. To test this, we substituted conserved residues in the Seb1 RRM that were shown for Nrd1 to be absolutely required for RNA binding (Bacikova *et al.*, 2014) (Fig. S5a). Substitutions in the RRM domain of Seb1 (RRM_{mut}) resulted in read through transcription (Fig. 4b) and affected poly(A) site selection (Fig. 4c, lane 5), similar to Seb1-deficient cells (lane 2). Moreover, we found that the CID and RRM mutants of Seb1 were affected in their recruitment at the 3' end of *rps2* (Fig. 4d, see region 2). In contrast, the Ser5_{mut} version of Seb1 showed only a modest reduction in cross-linking (Fig. 4d), consistent with the absence of defects in 3' end processing and transcription termination for this mutant. Importantly, the 3' end processing/transcription termination defects observed for the CID and RRM mutants were not the consequence of a problem in protein stability (Fig. S5b). The extent to which the different Seb1 mutants restored the growth defect induced by a Seb1 deficiency also correlated with the levels of 3' end processing/transcription termination defects (Fig. 4e). Taken together, these data argue that the function of Seb1 in 3' end

RNA processing and transcription termination is enforced by a specific mode of recruitment at the 3' end of genes that involves its CID and RRM domains.

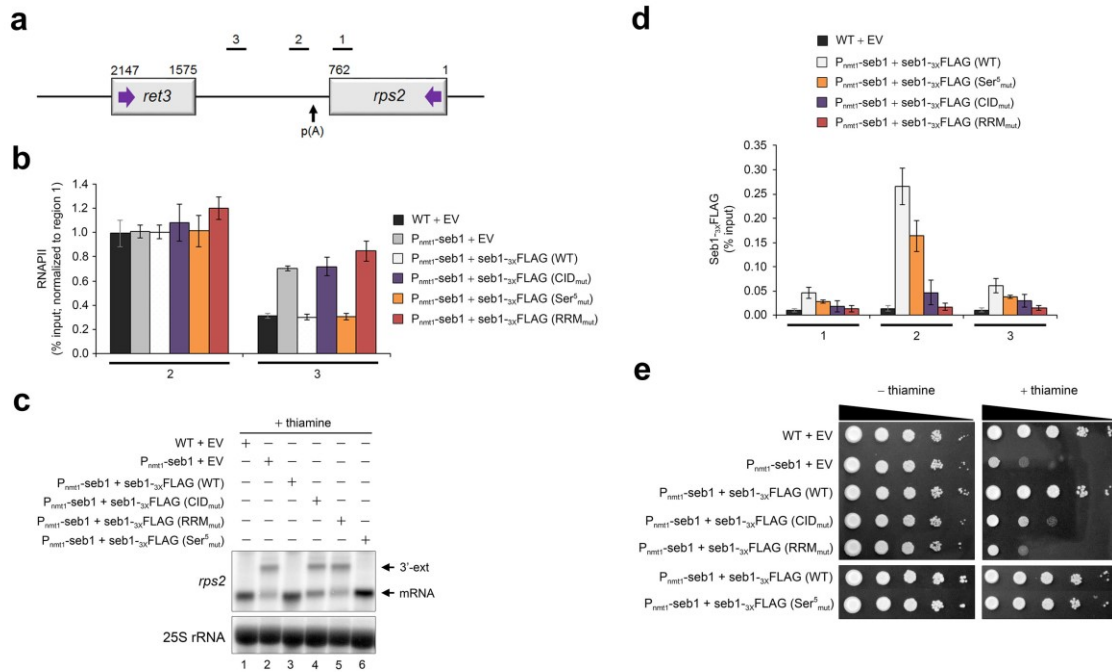


Figure 4. Seb1 requires functional CID and RRM domains for accurate 3' end processing and transcription termination.

(a) Bars above the *rps2* gene show the positions of PCR products used for ChIP analyses. (b) RNAPII ChIP analysis using extracts prepared from the P_{nmt1} -seb1 conditional strain containing genomically-integrated constructs that express the indicated versions of FLAG-tag Seb1: WT, CID_{mut}, Ser⁵_{mut}, and RRM_{mut} (see text and Fig. S5a for description), as well as an empty control vector (EV). Error bars, s.d. (n=3 biological replicates from independent cell cultures). (c) Northern blot analysis of *rps2* mRNA from the indicated strains. The *rps2* 3'-extended transcripts are shown (3'-ext). (d) ChIP analyses of wild-type and mutant versions of Seb1-FLAG along the *rps2* gene. Control wild-type cells with an empty vector (black bars) were used as negative control for the anti-FLAG ChIP assays. Error bars, s.d. (n=3 biological replicates from independent cell cultures). (e) Ten-fold serial dilutions of the indicated strains were spotted on thiamine-free (left) or thiamine-containing (right) minimal media.

Figure S5. Transcription termination defects upon mutations of the CID and RRM domains of Seb1.

(a) Sequences were aligned using Clustal Omega (Sievers et al. 2011). The RNAPII C-terminal-interacting domain (CID) and RNA recognition motif (RRM), as defined by the Simple Modular Architecture Research Tool (SMART) algorithm (Schultz et al. 1998), are boxed in red and blue, respectively. Consensus sequences involved in RNA interaction are identified as RNP1/RNP2 (Steinmetz and Brow. 1996). Conserved and similar residues are highlighted in black and gray, respectively. Function of the CID and the RRM domain of Seb1 were address using mutagenesis of conserved and/or similar residues. To create the phospho-Ser5 CID mutant (Ser⁵_{mut}), serine (S) residue 22 and lysine (K) residue 25 were substituted in combination to alanine residues (S22A/K25A) (blue-filled dots). The same approach was used for the other mutants generated. This include another mutant of the CID domain (CID_{mut}) where tyrosine (Y) residue 64, aspartic acid (D) residue 67 and arginine (R) residue 71 were all substituted to alanine (Y64A/D67A/R71A) (black-filled dots) as well as a RRM mutant (RRM_{mut}), where arginine (R) residue 442, histidine (H) residue 443, phenylalanine (F) residue 445 and lysine (K) residue 447 were all converted to alanine (R442A/H443A/F445A/K447A) (red-filled dots). (b) Western blot analysis of extracts prepared from P_{nmt1}-seb1 cells transformed with an empty vector or the indicated Seb1 constructs. (c) ChIP analyses of Seb1-3XFLAG association normalized to RNAPII density using extracts prepared from the indicated strains. Only the amplicon #2 is presented as it represents the chromatin region where Seb1 is preferentially associated. Errors bars in ChIP assays represent the standard deviation from 3 independent biological replicates or the range from two biological replicates in the case of the CID mutant strain. Experiments in (b) and (c) were performed in thiamine-containing medium.

Seb1 binds GUA-containing motifs downstream of cleavage sites

The functional requirement of the Seb1 RRM domain for the selection of 3' end cleavage sites prompted us to examine the position and sequence specificity implicated in the Seb1-RNA recognition. We used our functional HTP-tagged version of Seb1 to map RNA interactions of Seb1 at the transcriptome-wide level by UV-induced RNA-protein cross-linking and analysis of cDNA by high-throughput sequencing (CRAC) (Granneman *et al.*, 2009). Briefly, actively growing cells were UV-irradiated to forge covalent bonds between proteins and RNA, and subjected to affinity purification under stringent conditions to recover Seb1-associated RNAs (Fig. S6a-c), which were analyzed by high-throughput sequencing. Only a very small number of mappable reads were recovered from the untagged negative control samples, roughly 400-1000 times less compared to the Seb1 tagged samples. A breakdown of hits from 2 independent CRAC experiments revealed that nearly 70% of Seb1-

associated RNAs originate from protein-encoding genes and intergenic regions (Fig. 5a). Specifically, a genome-wide coverage plot showed that the majority of Seb1 RNA binding mapped downstream of genome annotations corresponding to open reading frames or noncoding RNAs (Fig. S6d), which is consistent with the genome-wide localization of Seb1 as determined by ChIP-seq. Indeed, a detailed analysis of the hit distribution across annotated poly(A) sites revealed preferential binding of Seb1 50-100 nucleotides downstream of cleavage sites (Fig. 5b, green plot & right axis), a profile that generally matched the distribution of mapped micro-deletions (Fig. 5b; red plot & left axis), which can be used to precisely map protein-RNA interactions (Granneman *et al.*, 2009)

Next, we searched for over-represented sequences in Seb1 read contigs using the pyMotif algorithm from the pyCRAC package (Webb *et al.*, 2014). Among the top scoring k-mers (4- to 8-mers) recovered from both CRAC experiments (Table S4), a clear GUA trinucleotide core was identified, which was surrounded by A/U and A/G as 5' and 3' nucleotides, respectively (Fig. 5c). Remarkably, the identified Seb1 consensus sequence is nearly identical to the GUA(A/G)-binding motif previously reported for Nrd1 (Bacikova *et al.*, 2014, Creamer *et al.*, 2011, Porrua *et al.*, 2012, Schaughency *et al.*, 2014, Schulz *et al.*, 2013, Wlotzka *et al.*, 2011), suggesting that sequence-specific recognition by the RRM domains of Seb1 and Nrd1 has been conserved despite divergent roles in RNA metabolism. Mapping the top Seb1 tetramers identified by CRAC (GUAG, UGUA; Table S4) along all fission yeast genes showed a strong enrichment downstream of cleavage sites, whereas gene bodies were markedly depleted of Seb1-binding motifs (Fig. 5d). In contrast, control tetramers with similar base composition, but lacking the GUA trinucleotide core, showed relatively even occurrence along genes (Fig. 5d).

To examine the functional significance of the identified Seb1 consensus motif, we analyzed 3' end processing of a model snoRNA gene, *snR99*, which shows 3'-extensions in Seb1-depleted cells (Fig. 5f, compare lanes 1-2). Accordingly, we used an *snR99*-null strain (Fig. 5f, lane 3) in which we re-introduced a wild-type copy of *snR99* or versions in which the putative Seb1 binding sites (see Fig. 5e) were deleted individually (mutant #1 or mutant #2) or in combination (mutant #1/2). Although the individual deletion of Seb1 consensus

motifs did not affect *snR99* 3' end processing (Fig. 5f; compare lanes 5-6 to lane 4), the combined deletion of two Seb1-binding motifs resulted in the accumulation of read-through *snR99* identical in length to those detected in Seb1-depleted cells (Fig. 5f; compare lanes 2 and 7). The fact that the double mutant, but not the single mutants, affected *snR99* 3' end processing suggests that multiple binding sites ensure Seb1 recruitment to promote accurate cleavage site position. Indeed, additional Seb1 consensus motifs can be found downstream of mutants #1/2 in the *snR99* 3' sequence (Fig. 5e), which likely explains the partial effect observed in the double mutant compared to the depletion of Seb1. Together, our results support a model in which Seb1 promotes polyadenylation site selection by binding onto nascent transcripts via a mechanism that relies on the recognition of GUA-containing motifs present downstream of 3' end cleavage sites.

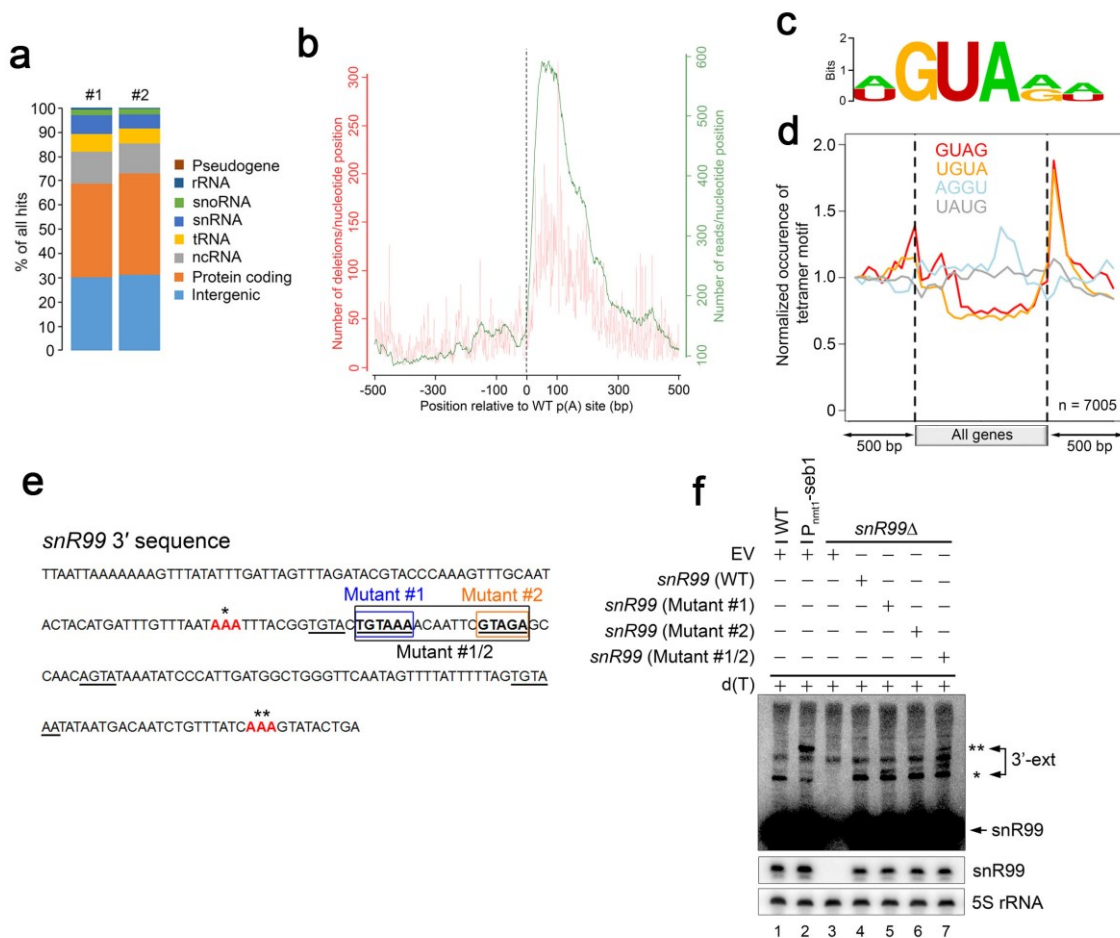


Figure 5. Seb1 binds to GUA-containing motifs downstream of p(A) sites.

(a) Distribution of Seb1-bound reads between transcript classes for two independent CRAC experiments. (b) Cumulative Seb1 RNA-binding sites relative to annotated p(A) sites. The green curve (right y-axis) shows the number of reads per nucleotide position, which is a measure of the binding preference. The red curve (left y-axis) shows the number of deletions per nucleotide position, which is an indication of direct crosslinking. (c) Sequence logo of Seb1 cross-linking sites derived from the WebLogo application (Crooks *et al.*, 2004) using the top 10 pyMotif-derived k-mers from each CRAC experiment. (d) Average gene distribution of tetrameric motifs derived from the Seb1 CRAC data (GUAG and UGUA) and control tetramers with shuffled di-nucleotides (AGGU and UAUG). (e) 200-nt of genomic sequence downstream of the *snR99* snoRNA gene with consensus motifs of the Seb1 binding site underlined. Regions indicated by colored boxes represent Seb1 binding sites that were deleted independently (blue box: Mutant #1; orange box: Mutant #2) or in combination (black box: Mutant #1/2). The red nucleotides indicate the major poly(A) site position of *snR99* polyadenylated precursors (3'-ext) detected in a wild-type (*) and in Seb1-depleted (**) cells, as determined by 3' RACE. (f) RNase H cleavage assay using *snR99*Δ cells (lanes 3-7) that contain an empty vector control (lane 3) or that express wild-type (lane 4) and mutant versions (lanes 5-7) of the *snR99* gene with deletions in the Seb1-binding motifs. WT (lane 1) and P_{nmt1}-seb1 (lane 2) cells transformed with an empty vector (EV) were analyzed as controls. Total RNA was treated with RNase H in the presence of a DNA oligonucleotide complementary to the *snR99* transcripts and an oligo(dT) to collapse the heterogeneous polyadenylated precursor population into a discrete product: *, normal 3' end processing site; **, 3' end processing site used in Seb1-deficient cells. Mature *snR99* snoRNA is indicated by an arrow.

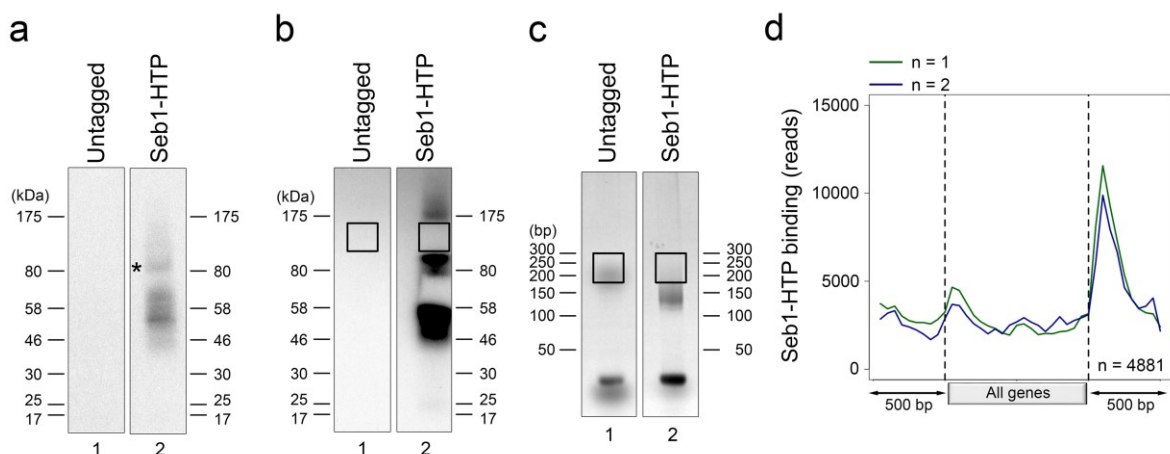


Figure S6. Mapping Seb1-HTP-bound RNAs by the CRAC technic.

(a-c) Purification of Seb1-RNA complexes. (a) Immunoblot analysis of Seb1-RNA cross-linked complexes purified via nickel affinity purification under denaturing conditions. *: full-length Seb1-HTP. (b) Autoradiogram of purified ³²P-labelled RNAs cross-linked to Seb1. Boxed regions were cut from the membrane and RNA isolated for cDNA libraries generation.

(c) Agarose gel electrophoresis of PCR products obtained from cDNA libraries. Boxed regions were gel extracted and submitted for high throughput sequencing. (d) Average gene distribution of preferential Seb1-HTP binding sites as determined by the analysis of Seb1-bound RNAs from two independent CRAC experiments. Light grey rectangles represent open reading frames or noncoding sequences and black arrows 500 base pairs (bp) of upstream and downstream flanking DNA sequences. The sample size is indicated at the bottom right of the graph. CRAC experiments were done in YES medium.

Seb1 promotes the recruitment of cleavage and polyadenylation factors

Given the copurification of Seb1 and cleavage/polyadenylation factors (Fig. 2), the binding of Seb1 50-100 nucleotides downstream of cleavage sites (Fig. 5), and its ability to favor proper poly(A) site selection (Fig. 3), we sought to investigate whether a deficiency in Seb1 influenced the recruitment of components of the 3' end processing machinery. We thus monitored the recruitment of 4 independent factors that belong to the CF1A (Rna14 and Clp1) and CPF (Ysh1 and Cft2) complex by ChIP assays. TAP-tagged versions of Rna14, Clp1, Ysh1, and Cft2 were all functional, as no growth phenotype was detected compared to the control untagged strain (Fig. S7a), and their expression were not affected by the depletion of Seb1 (Fig. S7b-e). These 3' end processing factors showed maximal recruitment near the poly(A) site of *rps2* in normal cells (Fig. 6a-6e, region 2), consistent with ChIP data from *S. cerevisiae* and human cells (Glover-Cutter *et al.*, 2008, Kim *et al.*, 2004, Licatalosi *et al.*, 2002). In contrast, this "bell-curve" distribution was lost for all tested proteins in Seb1-depleted cells, showing no accumulation around the poly(A) site, and accordingly, a flat distribution (Fig. 6a-e). Because *rps2*, and protein-coding genes in general, showed slightly lower levels of total RNAPII in Seb1-deficient cells (Fig. 1g-1i and Fig. S1a), we normalized the recruitment data to total polymerase levels, as measured by RNAPII ChIP signal performed with the same extracts. The normalized data revealed significant defects in Rna14 and Cft2 recruitment, whereas Clp1 and Ysh1 levels were similar between Seb1-deficient and control cells (Fig. 6f-i). Importantly, similar results were obtained for a snoRNA-encoding gene (Fig. S7f-l), indicating that this defective recruitment pattern is not limited to mRNA-encoding genes. We conclude that Seb1 functions in poly(A) site selection by promoting the co-transcriptional recruitment of components of the 3' end processing machinery.

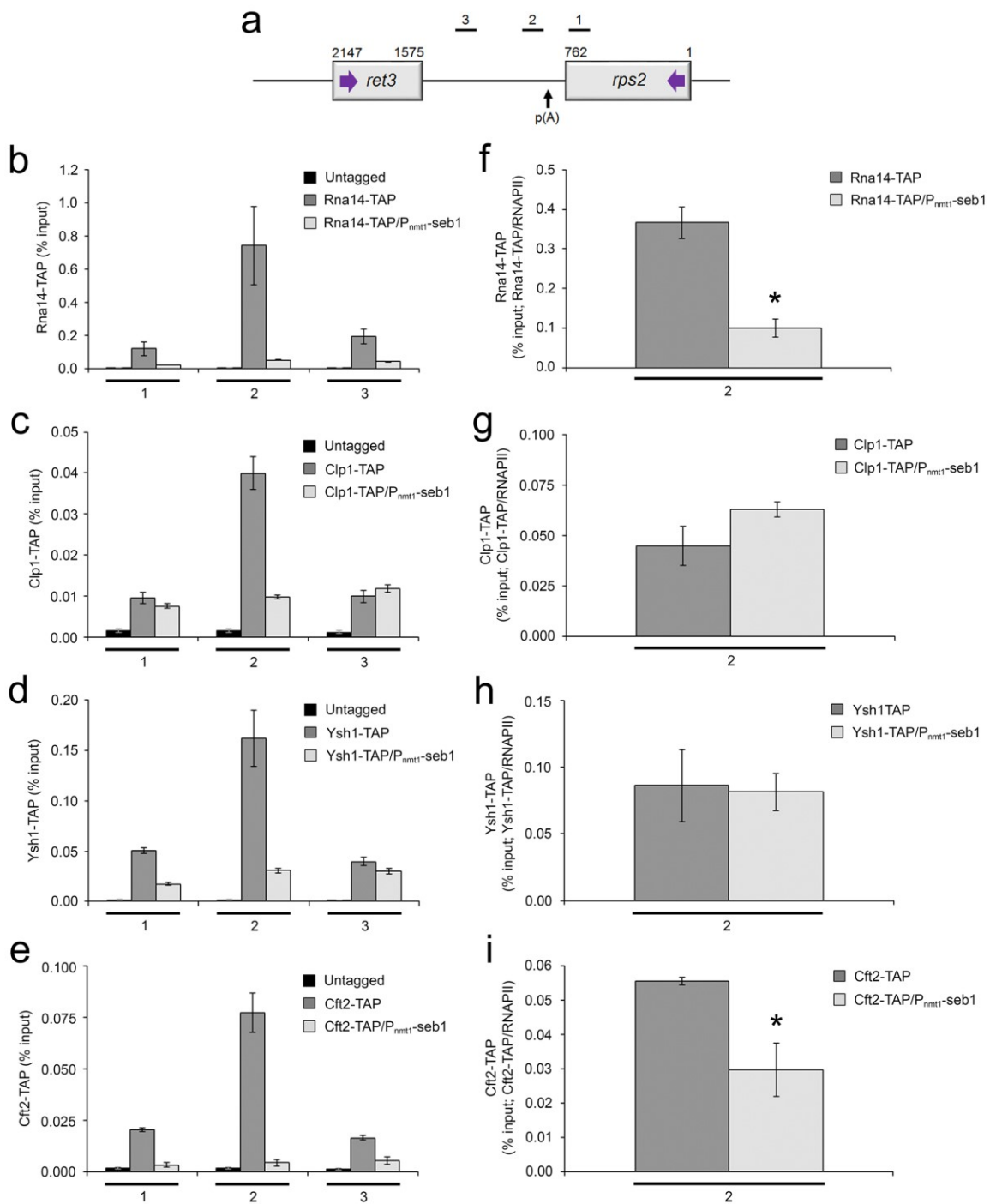


Figure 6. Seb1 levels affect the co-transcriptional assembly of the cleavage/polyadenylation machinery.

(a) Bars above the *rps2* gene show the positions of PCR products used for ChIP analyses. (b-e) ChIP assays of TAP-tagged versions of Rna14 (b), Clp1 (c), Ysh1 (d) and Cft2 (e) in wild-type and Seb1-depleted cells (*P_{nmt1}-seb1*). An untagged control strain was used to monitor the background signal of the ChIP assays. (f-i) Recruitment of 3' end processing factors as a ratio of total RNAPII at the 3' end of *rps2* (region 2). ChIP signal of TAP-tagged 3' end

processing factors were divided by the total RNAPII at region 2. Region 2 was analyzed because it represents the location of maximal 3' end processing factor recruitment (panels a-e). Error bars, s.d. (n=3 biological replicates from independent cell cultures). *: $p < 0.05$ (Student's t-test).

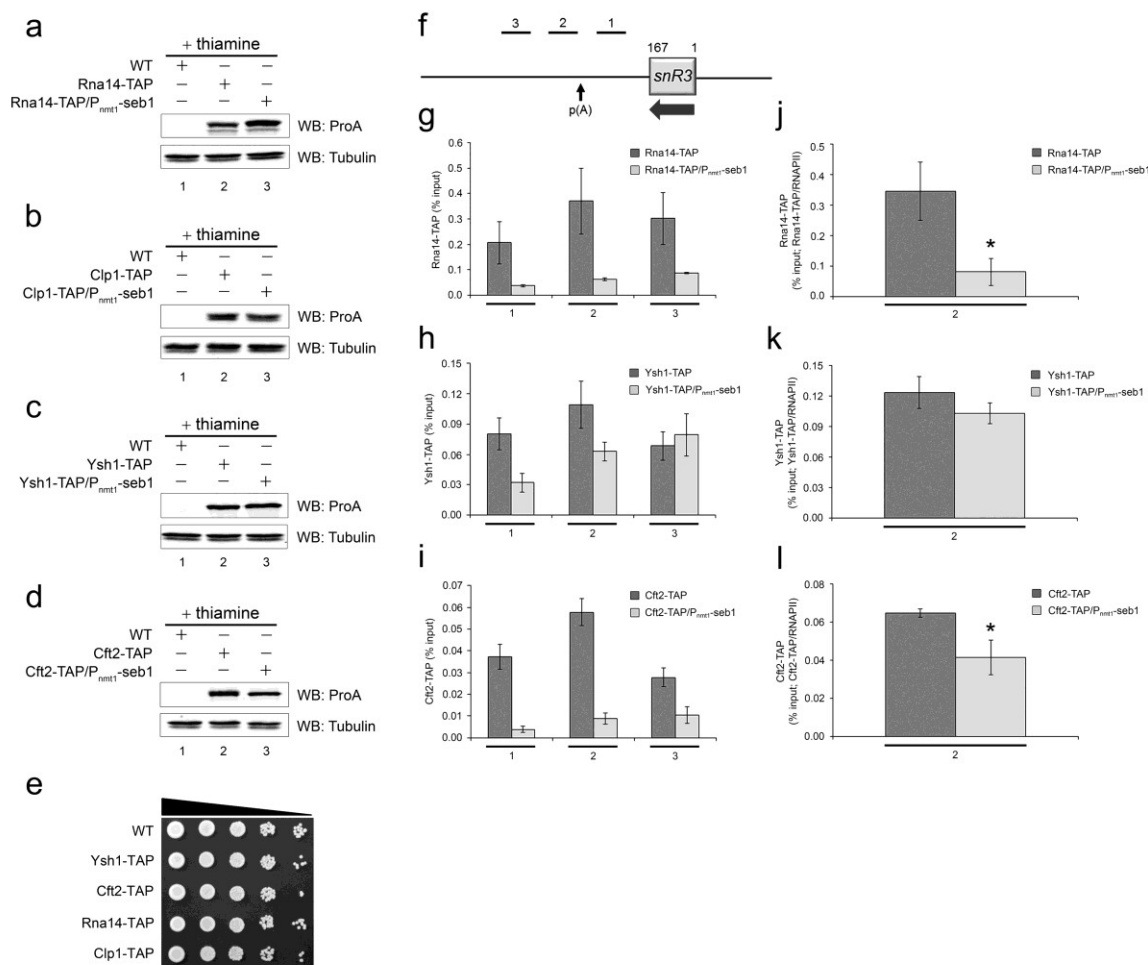


Figure S7. 3'-end processing factors are aberrantly recruited in Seb1-depleted condition.

(a-d) Immunoblot analyses of whole-cell extracts obtained from wild-type (WT) untagged cells or from WT and *P_{nmt1}-seb1* cells harboring a TAP-tag version of Rna14 (a), Clp1 (b), Ysh1 (c) or Cft2 (d). TAP-tagged proteins were revealed by an antibody directed against the protein A domain of the tag. Tubulin serves as a loading control. Experiments were conducted in thiamine-containing medium. (e) Ten-fold serial dilutions of the indicated strains were spotted on Yes medium. Rna14, Clp1, Ysh1, and Cft2 are all essential proteins (Kim et al. 2010). Yet, no growth apparent defect was observed arguing that the TAP-tagged proteins are functional. (f-l) (f) Schematic representation of the *snR3* locus. Bars above the gene show the positions of PCR products used for ChIP analyses throughout Figure S7. For *snR3*, p(A) refers to the polyadenylation sites of its 3'-extended precursors (Lemay et al. 2010). (g-i) ChIP assays on TAP-tagged versions of Rna14-TAP (g), Ysh1-TAP (h) and Cft2-TAP (i).

Cft2-TAP (i) in wild-type (WT) or Seb1-depleted condition (for example: Rna14-TAP versus Rna14-TAP/P_{nmt1}-seb1). (j-l) 3'-end processing factors recruitment normalized to RNAPII density. Only amplicon 2 is shown as it represents the region where 3'-end processing factors recruitment is maximal. Error bars represent the standard deviation from three independent biological replicates. *: $p < 0.05$ (Student's t-test). Experiments throughout Figure S7 were done in thiamine-supplemented medium.

Reduction of transcription elongation rates suppresses the cleavage site selection defects of Seb1-deficient cells.

Despite the absence of a global reduction in RNA levels (Fig. 3i), a general reduction in total RNAPII levels was observed in gene bodies in Seb1-deficient cells (Fig. 1f-i and Fig. S1). Such an observation may be indicative of a change in the transcription elongation rate. To determine if transcription kinetics contributes to poly(A) site selection by Seb1, we grew wild-type and Seb1-depleted cells in the presence of 6-azauracil (6-AU), which slows down transcription elongation and increases RNAPII pausing (Mason et Struhl, 2005). Treatment of wild-type cells with 6-AU resulted in a slight decrease in mRNA levels (Fig. 7a-b, compare lanes 1-2), consistent with reduced transcription kinetics. Notably, the addition of 6-AU to Seb1-deficient cells resulted in a marked change in the ratio between proximal and distal mRNA isoforms compared to non-treated cells. Specifically, 6-AU restored the altered polyadenylation site selection resulting from a Seb1 deficiency by reestablishing the levels of the normal, short isoform concurrent with reducing the levels of the distal isoform (Fig. 7a-b, compare lanes 3-4). These results show that slowing down transcription rates overcomes the need for Seb1 in selecting proper 3' end cleavage sites, suggesting an important role for RNAPII elongation kinetics in Seb1-dependent poly(A) site selection.

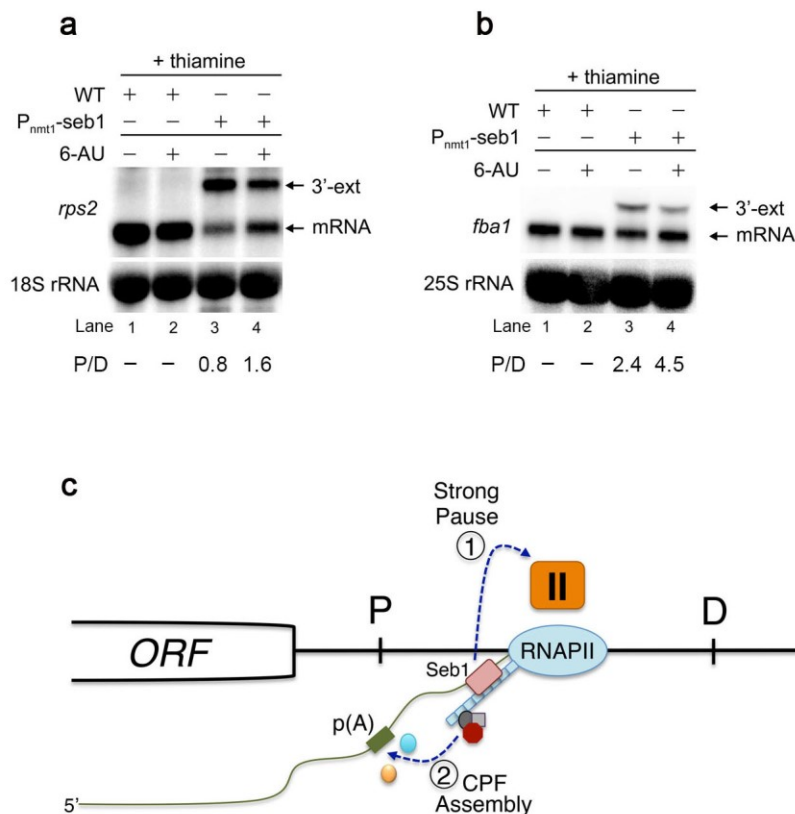


Figure 7. Transcription kinetics contributes to Seb1-dependent polyadenylation site selection.

(a-b) Northern blot analysis of total RNA prepared from wild-type (lanes 1-2) and Seb1-depleted cells (lanes 3-4) that were treated (lanes 2 and 4) or not treated (lanes 1 and 3) with 6-azauracil (6-AU). Blots were probed for *rps2* (a) and *fba1* (b) mRNAs. Ratios of proximal (P) relative to distal (D) mRNA isoforms are indicated (average from two independent experiments). (c) Model for Seb1-dependent poly(A) site selection. The passage of RNAPII through a poly(A) signal is thought to induce a change in the kinetics of transcription elongation, including pausing of the RNAPII complex (Beaudoin et Labbe, 2006, Davidson *et al.*, 2014, Grosso *et al.*, 2012, Nag *et al.*, 2006, Nojima *et al.*, 2015). We propose that the cooperative binding of Seb1 to the RNAPII CTD and to RNA motifs clustered downstream of poly(A) signals positively contribute to RNAPII pausing (1), thereby promoting poly(A) site recognition and assembly of a cleavage-competent cleavage/polyadenylation (CPF) complex (2). In the absence of Seb1, RNAPII pausing is leaky, increasing the frequency of RNAPII complexes that reach distal (D) poly(A) signals.

DISCUSSION

In this study, we identified an unexpected role for Seb1 in polyadenylation site selection and termination of RNAPII transcription. We also provide a framework into how Seb1 co-transcriptionally promotes accurate 3' end processing, thereby controlling the length of 3' untranslated regions (UTRs). These findings are significant, as they provide novel insights into the poorly understood mechanism that coordinates transcription and cleavage site selection, which underlie gene regulation by alternative polyadenylation.

NNS-like transcription termination is not evolutionarily conserved

In *S. cerevisiae*, transcription termination of noncoding genes does not depend on the mRNA 3' end processing machinery, but generally relies on a pathway that requires the NNS complex, which targets released transcripts to the RNA exosome for degradation or processing (Porrua et Libri, 2015). To date, however, the conservation of NNS-like transcription termination has remained elusive, as a distinctive NNS complex has not been described in metazoans. Instead, a complex consisting of the nuclear cap-binding proteins and ARS2 (CBCA complex) appears to share functional similarities to the *S. cerevisiae* NNS complex by connecting transcription of noncoding genes to exosome-dependent degradation in humans (Andersen *et al.*, 2013, Hallais *et al.*, 2013). Accordingly, the CBCA complex suppresses the production of read-through transcripts from several classes of noncoding genes by promoting transcription termination in a manner dependent on the distance from the transcription start site (Andersen *et al.*, 2013, Hallais *et al.*, 2013), which is similar to NNS-dependent termination in *S. cerevisiae* (Gudipati *et al.*, 2008, Vasiljeva *et al.*, 2008). In contrast to metazoans, *S. pombe* possesses putative orthologs of all NNS components: Seb1, Nab3, and Sen1/Dbp8. Yet, our proteomic analysis of Seb1 did not reveal the presence of a typical NNS-like complex. Instead, we found a significant enrichment of proteins involved in mRNA 3' end cleavage and polyadenylation. Such a physical connection with the 3' end processing machinery was not detected in proteomic analyses of Nrd1 (Vasiljeva et Buratowski, 2006), arguing that Seb1 and Nrd1 have functionally diverged. Indeed, our results indicate that neither Seb1 nor Nrd1 can functionally complement each other (Fig. S2).

Together with the absence of strong termination defects in cells deleted for *nab3*, *sen1*, and *dlb8* (Fig. 1), our findings argue that a machinery distinct from the budding yeast NNS complex connects noncoding RNA transcription to exosome-dependent RNA decay in fission yeast. Accordingly, a complex that is reminiscent to the human CBCA-NEXT complex has recently been described in *S. pombe*. This complex, known as MTREC or NURS (Egan *et al.*, 2014, Lee *et al.*, 2013, Zhou *et al.*, 2015), binds to noncoding, unwanted, and misprocessed transcripts and targets them for degradation by the nuclear exosome. Our study thus suggests that the NNS-exosome connection that functions in the recognition and degradation of aberrant transcripts in *S. cerevisiae* has functionally diverged in the *S. pombe* lineage, and has been lost over the course of metazoan evolution.

Seb1 controls polyadenylation site selection

Recent transcriptome-wide studies indicate that multiple polyadenylation sites (PAS) are used in most eukaryotic genes, as demonstrated in humans (Hoque *et al.*, 2013) as well as in budding (Rozenblatt-Rosen *et al.*, 2009) and fission (Mata, 2013, Schlackow *et al.*, 2013) yeasts. This process, known as alternative polyadenylation (APA), is emerging as a major layer of gene regulation, allowing the inclusion or exclusion of sequences that control the localization, stability, and translation of mRNAs (Tian et Manley, 2013). As yet, however, the mechanism of PAS recognition and how PAS selection is modulated remain poorly understood. Notably, our study disclosed a key role for the Seb1 RNA-binding protein in cleavage site selection *in vivo*, showing that a deficiency in Seb1 results in widespread changes in 3' UTR lengths as a consequence of increased APA. The direct role of Seb1 in PAS selection is supported by several observations: (i) Seb1 is specifically recruited at the 3' end of RNAPII-transcribed genes (Fig. 2 and Fig. S3), (ii) subunits of the 3' end processing machinery copurify with Seb1 (Fig. 2), (iii) Seb1 binds nascent transcripts 50-100 nt downstream of cleavage sites (Fig. 5), and (iv) RNA-seq data and western blotting analysis indicate that a Seb1 deficiency does not affect the expression of genes encoding for components of the 3' end processing machinery (Fig. S7 and data not shown). Our results therefore suggest that the heterogeneity of PAS usage in fission yeast (Mata, 2013,

Schlackow *et al.*, 2013) is not a purely random process, but can be modulated by controlling Seb1 levels.

Although Seb1 globally affected PAS selection, our standard RNA-seq data together with numerous northern blot validations did not reveal significant changes in mRNA abundance in Seb1-depleted cells. These observations argue that Seb1 is not required for the cleavage reaction itself, but rather to correctly position the cleavage site. To promote accurate 3' end processing, we found that Seb1 required a functional CID domain. A version of Seb1 with CID substitutions at conserved residues shown to be involved in CTD interactions in related CTD-binding proteins (Lunde *et al.*, 2010, Meinhart et Cramer, 2004) abolished Seb1 enrichment at the 3' end processing site and impaired polyadenylation site selection (Fig. 4). Our study also indicated that binding of Seb1 to (A/U)GUA(A/G)-containing motifs 50-100 nt downstream of PAS was important for Seb1 recruitment and 3' end cleavage site selection. Notably, this Seb1 binding signature downstream of PAS explains the reported lack of interaction between Seb1 and the *act1* mRNA (Marina *et al.*, 2013), as Seb1 is bound to the 3' fragment following cleavage instead of the mature polyadenylated mRNA. Collectively, our findings support a model where Seb1 is recruited to 3' end processing sites via interactions with the CTD of the RNAPII elongation complex and specific RNA motifs clustered downstream of PAS as they emerge from the transcribing polymerase. Such cooperative contribution of CID and RRM domains in the spatio-temporal recruitment of RNA processing factors is not uncommon during mRNA 3' end processing (Dichtl *et al.*, 2002, Kyburz *et al.*, 2003) and also underlies Nrd1-dependent recruitment (Gudipati *et al.*, 2008).

How does Seb1 promote the selection of polyadenylation sites? Our data indicated that Seb1 is important for the co-transcriptional organization of properly assembled cleavage and polyadenylation (CPF) complexes at the 3' end of genes (Fig. 6 and Fig. S7). We also showed that slowing down transcription rates and increasing RNAPII pausing frequency using 6-AU restored the cleavage site selection defects of Seb1-depleted cells (Fig. 7). Together, these findings suggest a model in which binding of Seb1 to clustered RNA motifs downstream of poly(A) signals modulates transcription elongation kinetics (RNAPII

pausing), enabling the recruitment and assembly of cleavage-competent CPF complexes (Fig. 7c). Defects in the co-transcriptional assembly of CPF complexes as a result of changes in transcription kinetics at the 3' end of genes may allow competition between polyadenylation sites as they emerge from the elongating polymerase, therefore providing a greater opportunity for the use of distal cleavage sites. A model in which Seb1-dependent poly(A) site selection is mechanistically linked to transcription elongation kinetics is supported by recent studies showing that RNAPII pausing influences the choice between alternative polyadenylation sites (Beaudoin et Labbe, 2006, Drissi *et al.*, 2015). Furthermore, definition of a minimal downstream element (DSE) important for RNAPII pausing in *S. pombe* (Aranda et Proudfoot, 1999) previously identified an 18-bp region that contains two copies of the pentanucleotide ATGTA, which is similar to the Seb1 RNA-binding motif determined by CRAC (Fig. 5c). A role for Seb1 in RNAPII pausing would also be consistent with proteomic analyses indicating that Seb1 is not a canonical subunit of the fission yeast 3' end processing machinery (Vanoosthuyse *et al.*, 2014), but a transient factor whose association is mediated by a paused transcription elongation complex (Fig. 7c).

Our ChIP analysis of Seb1-deficient cells detected a reduction in the occupancy of Rna14 and Cft2, which are subunits of the evolutionarily conserved CstF and CPSF complexes (Xiang *et al.*, 2014), respectively, at the 3' end of genes. However, Seb1 levels did not affect the recruitment of every component of CstF and CPSF complexes (Clp1 and Ysh1). Although the molecular basis underlying the specificity of 3' end factor recruitment by Seb1 remains to be determined, these data suggest that CstF and CPSF components may not be recruited to genes as complete pre-formed complexes, but may require stepwise assembly processes that occur co-transcriptionally, which is consistent with previous work (Chao *et al.*, 1999, Lackford *et al.*, 2014, Martin *et al.*, 2012). Consistent with this idea, a complex containing Cft2 and other CPF components can be isolated independently of Ysh1 in *S. cerevisiae* (Cox et Mann, 2008). In addition, the observation that Seb1 levels did not impair the overall recruitment of Ysh1, which is the fission yeast homolog of the human endonuclease CPSF73, is also consistent with the conclusion that Seb1 is not essential for the cleavage reaction itself, but necessary to correctly position the site of cleavage.

Remarkably, the absence of an NNS-like transcription termination pathway in fission yeast argues for the involvement of the mRNA cleavage and polyadenylation machinery in *S. pombe* snoRNA 3' end processing, which contrasts to snoRNA 3' end formation in *S. cerevisiae* that relies on the NNS complex (Porrua et Libri, 2015). Accordingly, we found that Seb1 and mRNA 3' end processing factors were generally recruited at the 3' end of both mRNA and snoRNA genes. We also observed termination defects and altered recruitment of mRNA 3' end processing factors at both mRNA and snoRNA genes in Seb1-depleted cells. Consistent with this idea, snoRNAs display mRNA-like features in fission yeast. For instance, mature snoRNAs are produced from polyadenylated precursors as part of their maturation cycle, which requires the activity of the canonical poly(A) polymerase (Pla1) and the nuclear poly(A)-binding protein Pab2 (Lemay *et al.*, 2010). Moreover, recent genome-wide mapping of poly(A) sites in fission yeast revealed that the most prevalent *cis* element associated with cleavage sites identification are common between mRNA and snoRNA genes (Mata, 2013, Schlackow *et al.*, 2013).

The importance of precisely selecting the correct poly(A) site is key during embryonic development and is of primordial importance for human health (Curinha *et al.*, 2014). The identification of Seb1 as an essential factor that can influence 3' end processing decision in a co-transcriptional manner is an important advance in understanding the interplay between transcription and APA regulation. Given the similarities between *S. pombe* and human polyadenylation signals (Mata, 2013, Schlackow *et al.*, 2013), including use of the canonical AAUAAA hexamer, together with the fact that metazoans include several proteins that possess both CID and RRM domains (Cox et Mann, 2008), we predict that the links between transcription and polyadenylation site selection described in fission yeast are likely to apply in higher eukaryotes.

EXPERIMENTAL PROCEDURES

Yeast strains and media

A list of all strains (*S. pombe* and *S. cerevisiae*) and plasmids used in this study is provided in Table S5-7. Fission yeast cells were grown at 30°C in yeast extract medium with adenine, uracil and amino acid supplements (YES) or in Edinburgh minimal media (EMM) supplemented with adenine, uracil and appropriated amino acids.

Chromatin immunoprecipitation (ChIP) assays

ChIP-qPCR and ChIP-seq experiments were performed as described previously (Lemay *et al.*, 2014). The antibodies used are described in Supplemental Experimental Procedures.

Protein analyses

Analysis of protein expression and affinity purification methods are described in details in the Supplemental Experimental Procedures.

3' READS analysis

The 3'READS method used in this study was performed and analyzed as previously described using total *S. pombe* RNA (Hoque *et al.*, 2013).

CRAC assays

CRAC was performed as previously described (Granneman *et al.*, 2009) using *S. pombe* cells were grown in YES medium to an OD_{600nm} of 0.45-0.5 and UV-irradiated in the Megatron UV cross-linker for 220 seconds (Granneman *et al.* 2011). Additional details can be found in the Supplementary Experimental Procedures.

Computational Methods

Reads obtained from Illumina HiSeq runs were quality filtered according to the Illumina pipeline. Detailed computational methods for RNA-seq, ChIP-seq, and CRAC analyses are described in the Supplementary Experimental Procedures.

ACKNOWLEDGMENTS

We thank James Manley for critical reading of the manuscript; D. Hermand, D. Libri, and R. Wellinger for strains, plasmids, and reagents; M. Durant and E. Lapointe (Laboratoire de Génomique Fonctionnelle de l'Université de Sherbrooke) for ChIP-seq libraries; S.R. Atkinson (University College London) for RNA-seq libraries; the sequencing platforms of the McGill University and Génome Québec Innovation Centre and Edinburgh Genomics. This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) to F.B., from a Wellcome Trust Senior Investigator Award to J.B., by the UK Medical Research Council to S.M.; by funding (GM084089) from the National Institute of General Medical Sciences to B.T.; and by a Wellcome Trust Research and Career Development Grant (091549) to S.G. F.B. is supported as a Canada Research Chair in Quality of Gene Expression.

REFERENCES

- Andersen PR, Domanski M, Kristiansen MS, Storvall H, Ntini E, Verheggen C, Schein A, Bunkenborg J, Poser I, Hallais M et al. 2013. The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol* 20: 1367-1376.
- Aranda A, Proudfoot NJ. 1999. Definition of transcriptional pause elements in fission yeast. *Molecular and cellular biology* 19: 1251-1261.
- Arigo JT, Eyler DE, Carroll KL, Corden JL. 2006. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular cell* 23: 841-851.
- Bacikova V, Pasulka J, Kubicek K, Stefl R. 2014. Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic acids research* 42: 8024-8038.

- Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. *Nature reviews Genetics* 15: 163-175.
- Berriz GF, Beaver JE, Cenik C, Tasan M, Roth FP. 2009. Next generation software for functional trend analysis. *Bioinformatics* 25: 3043-3044.
- Chao LC, Jamil A, Kim SJ, Huang L, Martinson HG. 1999. Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. *Molecular and cellular biology* 19: 5588-5600.
- Corden JL. 2013. RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chemical reviews* 113: 8423-8455.
- Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ, Corden JL. 2011. Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet* 7: e1002329.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14: 1188-1190.
- Curinha A, Oliveira Braz S, Pereira-Castro I, Cruz A, Moreira A. 2014. Implications of polyadenylation in health and disease. *Nucleus* 5: 508-519.
- Davidson L, Muniz L, West S. 2014. 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev* 28: 342-356.
- Dichtl B, Blank D, Sadowski M, Hubner W, Weiser S, Keller W. 2002. Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J* 21: 4125-4135.
- Egan ED, Braun CR, Gygi SP, Moazed D. 2014. Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *RNA* 20: 867-881.
- Fusby B, Kim S, Erickson B, Kim H, Peterson ML, Bentley DL. 2015. Coordination of RNA Polymerase II Pausing and 3' End Processing Factor Recruitment with Alternative Polyadenylation. *Molecular and cellular biology* 36: 295-303.
- Ghazy MA, Gordon JM, Lee SD, Singh BN, Bohm A, Hampsey M, Moore C. 2012. The interaction of Pcf11 and Clp1 is needed for mRNA 3'-end formation and is modulated by amino acids in the ATP-binding site. *Nucleic acids research* 40: 1214-1225.

- Glover-Cutter K, Kim S, Espinosa J, Bentley DL. 2008. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol* 15: 71-78.
- Granneman S, Kudla G, Petfalski E, Tollervey D. 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A* 106: 9613-9618.
- Granneman S, Petfalski E, Tollervey D. 2011. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *EMBO J* 30: 4006-4019.
- Grosso AR, de Almeida SF, Braga J, Carmo-Fonseca M. 2012. Dynamic transitions in RNA polymerase II density profiles during transcription termination. *Genome Res* 22: 1447-1456.
- Gudipati RK, Villa T, Boulay J, Libri D. 2008. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol* 15: 786-794.
- Hallais M, Pontvianne F, Andersen PR, Clerici M, Lener D, Benbahouche Nel H, Gostan T, Vandermoere F, Robert MC, Cusack S et al. 2013. CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. *Nat Struct Mol Biol* 20: 1358-1366.
- Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nat Methods* 10: 133-139.
- Jensen TH, Jacquier A, Libri D. 2013. Dealing with pervasive transcription. *Molecular cell* 52: 473-484.
- Johnson SA, Cubberley G, Bentley DL. 2009. Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. *Molecular cell* 33: 215-226.
- Kim M, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S. 2004a. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *Embo J* 23: 354-364.

- Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedeja E, Greenblatt JF, Buratowski S. 2004b. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432: 517-522.
- Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, Hofr C, Vanacova S, Stefl R. 2012. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26: 1891-1896.
- Kyburz A, Sadowski M, Dichtl B, Keller W. 2003. The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. *Nucleic acids research* 31: 3936-3945.
- Lee NN, Chalamcharla VR, Reyes-Turcu F, Mehta S, Zofall M, Balachandran V, Dhakshnamoorthy J, Taneja N, Yamanaka S, Zhou M et al. 2013. Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* 155: 1061-1074.
- Lemay JF, D'Amours A, Lemieux C, Lackner DH, St-Sauveur VG, Bahler J, Bachand F. 2010. The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. *Molecular cell* 37: 34-45.
- Lemay JF, Larochelle M, Marguerat S, Atkinson S, Bahler J, Bachand F. 2014. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat Struct Mol Biol* 21: 919-926.
- Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL. 2002. Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Molecular cell* 9: 1101-1111.
- Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17: 1195-1201.
- Marina DB, Shankar S, Natarajan P, Finn KJ, Madhani HD. 2013. A conserved ncRNA-binding protein recruits silencing factors to heterochromatin through an RNAi-independent mechanism. *Genes Dev* 27: 1851-1856.
- Mason PB, Struhl K. 2005. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell* 17: 831-840.

- Mata J. 2013. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol* 10: 1407-1414.
- Mayer A, Heidemann M, Lidschreiber M, Schrieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P. 2012. CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336: 1723-1725.
- Meinhart A, Cramer P. 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430: 223-226.
- Nag A, Narsinh K, Kazerouninia A, Martinson HG. 2006. The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. *RNA* 12: 1534-1544.
- Nojima T, Gomes T, Grosso AR, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* 161: 526-540.
- Oktaba K, Zhang W, Lotz TS, Jun DJ, Lemke SB, Ng SP, Esposito E, Levine M, Hilgers V. 2015. ELAV links paused Pol II to alternative polyadenylation in the *Drosophila* nervous system. *Molecular cell* 57: 341-348.
- Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM. 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* 143: 1018-1029.
- Porrúa O, Hobor F, Boulay J, Kubicek K, D'Aubenton-Carafa Y, Gudipati RK, Stefl R, Libri D. 2012. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J* 31: 3935-3948.
- Porrúa O, Libri D. 2013. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol* 20: 884-891.
- Porrúa O, Libri D. 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* 16: 190-202.
- Schaughency P, Merran J, Corden JL. 2014. Genome-wide mapping of yeast RNA polymerase II termination. *PLoS Genet* 10: e1004632.
- Schlackow M, Marguerat S, Proudfoot NJ, Bahler J, Erban R, Gullerova M. 2013. Genome-wide analysis of poly(A) site selection in *Schizosaccharomyces pombe*. *RNA* 19: 1617-1631.

- Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J, Cramer P. 2013. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* 155: 1075-1087.
- Shi Y, Manley JL. 2015. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev* 29: 889-897.
- Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, Libri D. 2006. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the *nrd1-nab3* pathway in genome surveillance. *Mol Cell* 23: 853-864.
- Tian B, Manley JL. 2013. Alternative cleavage and polyadenylation: the long and short of it. *Trends Biochem Sci* 38: 312-320.
- Vanoosthuysse V, Legros P, van der Sar SJ, Yvert G, Toda K, Le Bihan T, Watanabe Y, Hardwick K, Bernard P. 2014. CPF-associated phosphatase activity opposes condensin-mediated chromosome condensation. *PLoS Genet* 10: e1004415.
- Vasiljeva L, Buratowski S. 2006. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Molecular cell* 21: 239-248.
- Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. 2008a. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 15: 795-804.
- Vasiljeva L, Kim M, Terzi N, Soares LM, Buratowski S. 2008b. Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. *Mol Cell* 29: 313-323.
- Webb S, Hector RD, Kudla G, Granneman S. 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* 15: R8.
- Wlotzka W, Kudla G, Granneman S, Tollervey D. 2011. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J* 30: 1790-1803.
- Xiang K, Tong L, Manley JL. 2014. Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. *Molecular and cellular biology* 34: 1894-1910.
- Zhou Y, Zhu J, Schermann G, Ohle C, Bendrin K, Sugioka-Sugiyama R, Sugiyama T, Fischer T. 2015. The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun* 6: 7050.

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

Yeast strains and media

Conditional strain in which the genomic copy of *seb1* is expressed from the thiamine-repressible *nmt1* promoter (Pnmt1) was generated as described previously (Granneman *et al.*, 2009). Pnmt1-dependent gene expression was repressed by the addition of 60 μ M of thiamine to the growth medium for 10-15 hrs. To inhibit transcription elongation, 6-azauracil (6-AU) was added to the culture medium of FBY1712 and FBY2063 strains at a final concentration 0.08 mg/ml for 4h. Gene disruptions and C-terminal tagging of proteins were performed by PCR-mediated gene targeting (Bähler *et al.* 1998), using the lithium acetate method for cell transformation. Knockouts strains were confirmed by RT-PCR, tagged proteins by western blotting and promoter swap by PCR screening and/or growth assay. Growth of *S. cerevisiae* strains were carried out in Synthetic Defined media (SD) supplemented with 2 % galactose, adenine, uracil and appropriated amino acids with the exception of leucine. Metabolic depletion of Nrd1p from the *P_{GALI-3XHA}-NRD1* strain was performed for 7 hrs in 2 % glucose-containing medium as described previously (Thiebaut *et al.* 2006). Primer sequences used throughout this study are available upon request.

Growth (spot) assays

For growth assays, exponential cell cultures were adjusted to an optical density (OD_{600nm}) of 1.0, and serially diluted 10-fold with water. For *S. pombe* experiments, each sample was spotted (3 μ l/spot) on EMM plates with or without 15 μ M of thiamine and incubated at 30°C for 3 and 7 days respectively. For *S. cerevisiae* spot assays, each sample was spotted (3 μ l/spot) on SD leu- plates containing either 2% galactose or glucose and incubated at 30°C for 3 and 7 days respectively.

RNA preparation and analyses

Total yeast RNA was extracted using the hot-acid phenol method, resolved on agarose formaldehyde gels or on polyacrylamide-urea gels in case of RNase H experiments and subsequently transferred onto nylon membranes prior to RNAs crosslinking. Pre hybridization and overnight hybridization were carried out in Church buffer at 42°C and 65°C for DNA and RNA probes, respectively. Strand-specific RNA probes were designed to be complementary to ORF sequences. RNA probes were generated by *in vitro* transcription using the T7 riboprobe system (NEB) and internally labeled with [α - ^{32}P]-UTP. DNA probes were generated by 5' end labeling of single-stranded oligonucleotides using the T4 polynucleotide kinase and [γ - ^{32}P]-ATP. For DNA probes, 2× 15 min washes were done in 2× SSC/0.1 % SDS followed by 2× 5 min washes in 0.1× SSC/0.1 % SDS; for RNA probes, 3× 15 min washes were performed in 1× SSC/0.1 % SDS followed by one 15 min wash in 0.1× SSC/0.1 % SDS. Membranes were exposed using Phosphor Screens followed by visualization and quantification of signals using a Typhoon Trio instrument and ImageQuant TL, respectively (GE Healthcare).

RNase H cleavage assays

RNase H assays were performed as described previously (Lemay et al. 2010).

Protein analyses

Analysis of protein expression. Total cell extracts were prepared by harvesting mid-log cells in ice-cold lysis buffer (50 mM Tris (pH 7.5), 5 mM MgCl₂, 150 mM NaCl and 0.1 % NP-40) containing a cocktail of protease inhibitors (1× PMSF, 1× PLAAC and 1× cOmplete) prior to lysis with glass beads using a FastPrep instrument (MP Biomedicals). Clarified lysates were normalized for total protein concentration using the Bradford protein assay. Proteins were separated by SDS-PAGE, transferred to nitrocellulose membranes, and analyzed by immunoblotting using either, a mouse monoclonal antibody specific to the FLAG peptide (Sigma-Aldrich, F1804; 1:500 (v/v) dilution), a mouse monoclonal antibody

specific to α -tubulin (Sigma-Aldrich, T5168; 1:1,000 (v/v) dilution), a mouse monoclonal antibody against the hemagglutinin protein (HA) (Roche, 11 583 816 001; 1:500 (v/v) dilution) or a rabbit polyclonal antibody against protein A (Sigma-Aldrich, P3775; 1:10,000 (v/v) dilution) in the case of HTP-tagged Seb1 or TAP-tagged proteins. Membranes were then probed with goat anti-rabbit or anti-mouse secondary antibodies conjugated to IRDye 800CW (LI-COR, 926-32213; 1:15,000 (v/v) dilution) and AlexaFluor 680 (Life Technologies, A-21057; 1:15,000 (v/v) dilution), respectively. Detection of the proteins was performed using an Odyssey infrared imaging system (LI-COR).

Small-scale affinity purification. For standard co-immunoprecipitation experiments, 10 mg of total proteins were subjected to immunoprecipitation with 25 μ l of Pan Mouse IgG Dynabeads (ThermoFisher Scientific, 11041), that were pre-equilibrated in ice-cold lysis buffer (50 mM HEPES-KOH (pH 7.5), 5 mM MgCl₂, 100 mM KCl, 0.25 % NP-40, 10 % glycerol, 1 mM EDTA and 0.5 mM DTT) containing a cocktail of phosphatase and protease inhibitors (1 \times PhosSTOP, 1 \times PMSF, 1 \times PLAAC and 1 \times cOmplete), and gently mixed for 2 hrs. When needed, 1 μ l of the DNA/RNA nuclease Benzonase (Sigma-Aldrich; E1014) (final concentration: 250 U/ml) was added to the mix at the beginning of the immunoprecipitation. Co-immunoprecipitated proteins were eluted by heat (95°C/5 min) in 1 \times SDS sample buffer. Proteins were separated by SDS-PAGE, transferred to nitrocellulose membranes, and analyzed by immunoblotting using, a mouse monoclonal antibody specific to the MYC peptide (Santa Cruz Biotechnology, sc-40; 1:500 (v/v) dilution) and a rabbit polyclonal antibody against protein A. Secondary antibodies and protein detection are as described above.

Large-scale affinity purification and mass spectrometry. A 2 L culture of WT and Seb1-HTP cells (OD_{600nm} of 1) grown in YES media was pelleted and resuspended in a volume of lysis buffer (50 mM HEPES-KOH (pH 7.5), 5 mM MgCl₂, 100 mM KCl, 0.25 % NP-40, 10 % glycerol, 1 mM EDTA, 0.5 mM DTT, 1 \times PhosSTOP, 1 \times mM PMSF, 1 \times PLAAC and 1 \times cOmplete) equal to one-fourth of the cell pellet volume and subsequently added dropwise to liquid nitrogen. Cells were disrupted by cryo-grinding with a Freezer/Mill 6870 (SPEX SamplePrep) using the following setup: 4 rounds of 2 min of grinding at 15

pulses/second followed by 2 min of cooling. Cell powder was resuspended in a volume of IP buffer (lysis buffer without DTT) equal to 1.2× the cell pellet volume and stirred for 15 min in a cold room. The lysate was then clarified by two successive centrifugations at 12,000 rpm at 4°C (25 and 15 min for the first and second centrifugation respectively). The supernatant of each sample was incubated for 2 hrs at 4°C on a rocker with 275 µl of Pan Mouse IgG Dynabeads that were pre-equilibrated in IP buffer. Beads were washed 3× with 1 ml of IP buffer for 1 min each followed by one 5 min wash. Proteins were eluted with freshly prepared 0.5 M NH₄OH at 37°C for 20 min and samples were dried overnight using a SpeedVac concentrator. Immunoprecipitated proteins were resuspended in 45 µl of buffer composed of 50 mM Tris (pH 6.8), 1 % SDS and 10 mM DTT and allow to reduce at 56°C for 30 min. Upon reaching room temperature, alkylating agent iodoacetamide was added to a final concentration of 50 mM and samples were incubated for 40 min in the dark prior to boiling in SDS sample buffer. The eluates were ultimately run on a 4-12 % Bis-Tris Novex mini-gel (Life Technologies) using MOPS buffer followed by Coomassie-staining (Simply Blue SafeStain; ThermoFisher Scientific). The entire protein gel lanes were excised and cut into 8 slices. Gel bands were destained and subjected to in-gel digestion with 12.5 µg/ml of freshly prepared trypsin Gold (Promega, V5280) overnight at 30°C. Digested peptides were extracted from gels bands, dried using a SpeedVac concentrator and resuspended in 0.1 % TFA (trifluoroacetic acid). The samples were then loaded onto mini-C18 columns (ZipTips; EMD Millipore) and process according to manufacturer's instructions but with slight modifications: the wetting and elution solution being respectively changed to 100 % acetonitrile and 1 % FA (formic acid)/50 % acetonitrile. Eluted peptides were dried using a SpeedVac concentrator and resuspended in 1 % FA. Purified trypsin digested samples were analyzed by liquid-chromatography (LC)-MS/MS as described previously (Grenier St-Sauveur et al. 2013).

***seb1*/NRD1 expression constructs**

seb1 expression constructs were created by a 3-step cloning procedure using the *ade6* integration vector 44, pFB366, as the host vector and contain 1 kbp of *seb1* 5' UTR sequences followed by the *seb1* open reading frame (ORF) and an additional 1 kbp of *seb1* 3' UTR

sequences. For complementation studies, the ORF of *nrd1* was cloned between *seb1*'s 5' and 3' UTR sequences. Insertion of a C-terminal 3xFLAG tag to both Nrd1p and Seb1 was performed using the Q5 site-directed mutagenesis kit (NEB) with primers containing the tag sequence resulting in pFB941 and pFB943 respectively. To create the C-terminal interacting domain (CID) mutant (CID_{mut}) of Seb1 (pFB944), tyrosine (Y) residue 64, aspartic acid (D) residue 67 and arginine (R) residue 71 were all changed to alanine (Y64A/D67A/R71A). For the Ser5 mutant (Ser5_{mut}), serine (S) residue 22 and lysine (K) residue 25 were both substituted to alanine (S22A/K25A = pFB1037). All mutations were performed by site-directed mutagenesis using pFB943 as template. The same approach was used for the RNA recognition motif (RRM) mutant (RRM_{mut}) (pFB946), where arginine (R) residue 442, histidine (H) residue 443, phenylalanine (F) residue 445 and lysine (K) residue 447 were all changed to alanine (R442A/H443A/F445A/K447A). All constructs were confirmed by sequencing. For single integration into the *ade6* locus, pFB366, pFB941, pFB943, pFB946 and pFB1037 were linearized and transformed into the conditional Pnmt1-*seb1* strain (FBY915). Positive integrants were confirmed by western blot. For complementation studies in *S. cerevisiae*, *NRD1-3xFLAG* and *SEB1-3xFLAG* cDNAs were cloned between 1 kbp of *NRD1*'s 5' and 3' sequence using pRS315 (pFB950) as the host vector resulting in pFB970 and pFB1007 respectively. All constructs were confirmed by sequencing. pFB950, pFB974 and pFB1007 were maintained episomally into the conditional *P_{GAL1-3xHA-NRD1}* strain (FBY1802) (Thiebaut et al. 2006). Expressions were confirmed by western blot.

***snR99* construct with alteration in Seb1-binding elements**

The *snR99* mutants designed to alter the binding of Seb1 to nascent *snR99* transcripts (*snR99* Mutant #1, #2 and #1/2) were generated using the *ade6* integration vector (pFB366) as the host vector and contain 1 kbp of *snR99* 5' sequences followed by the *snR99* noncoding region and an additional 1 kbp of *snR99* 3' sequences, to result in pFB600. Deletion of nucleotides 91-96 (inclusive) (Mutant #1), 104-108 (Mutant #2) and 91-108 (Mutant #1/2) of the *snR99* 3' sequences were performed using the Q5 site-directed mutagenesis kit with pFB600 as template to result in pFB1057, pFB1059 and pFB1061 respectively. All constructs were confirmed by DNA sequencing. For single integration into the *ade6* locus, pFB600,

pFB1057, pFB1059 and pFB1061 were linearized and transformed into a *snR99* Δ strain. Positive integrants were confirmed by growth selection on EMM agar plates lacking adenine while similar *snR99* transcript level between the WT and mutated versions of *snR99* were confirmed by rt-PCR using primers located within the *snR99* noncoding region.

Chromatin immunoprecipitation (ChIP) assays

In the case of *Pnmt1-seb1* cells grown in thiamine-supplemented condition, the volume of cell culture was doubled to get sufficient RNAPII-bound chromatin for library generation. Chromatin was immunoprecipitated directly with Pan Mouse IgG Dynabeads in the case of TAP/HTP-tagged proteins or with anti-FLAG magnetic agarose beads (Sigma-Aldrich, M8823) when the *Seb1-₃XFLAG* strain and its CID, Ser5 and RRM mutants versions were used. ChIP assays were also performed using Pan Mouse IgG Dynabeads coated with antibodies against Rpb1 (8WG16; Convance, MMS-126R) or the hemagglutinin protein (HA) when using the *Rpb3-3XHA* strain. Note that the 8WG16 antibody is known to recognize the C-terminal domain (CTD) of Rpb1 and measures total RNAPII along genes in yeast (Bataille et al. 2012). Control ChIP assays with untagged strains or with isotype matched control antibody were performed.

Transcription run-on assays

Transcription run-on assays were performed as described previously (Lemay et al. 2014).

Library preparation and Illumina sequencing

RNA-seq Libraries were prepared from total *S. pombe* RNA and sequenced with Illumina HiSeq technology as described (Lemieux et al. 2009). For ChIP-seq analyses, libraries were prepared as described previously (Lemay et al. 2014).

RNA-seq data analysis. RNA-seq data were analysed as in (Lemay et al. 2014).

ChIP-seq data analysis. Data were analysed as described in (Lemay et al. 2014). Average gene and reads accumulation analyses were performed with in house R scripts using TSS coordinates generated in this study (poly(A) mapping). TSS coordinates were retrieved from the Pombase database (www.pombase.org). Heatmaps of read coverage were generated using the deepTools software (Ramirez et al. 2014) using TSS and TTS coordinates reported in the Pombase database. Single gene profiles were generated using the Integrative Genomics Viewer (Robinson et al. 2011).

CRAC assays, libraries preparation and Illumina sequencing

CRAC was performed as previously described (Granneman et al. 2009) with slight modifications: we used App_PE 3' adapter (5'-App-NAGATCGGAAGAGCACACGTCTG-ddC) and App_RT (5'-CAGACGTGTGCTCTTCCGATCT) oligonucleotides to make the libraries compatible with paired-end sequencing. The cDNA libraries were PCR amplified using P5 (5'-AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCT) and barcoded P3 oligonucleotides (5'-CAAGCAGAAGACGGCATAACGAGAT-6nt_barcode-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT) and CRAC libraries were paired-end sequenced on a HiSeq2500 at Edinburgh Genomics, University of Edinburgh. Data from two biological replicates (untagged control and Seb1-HTP samples) were subsequently merged.

CRAC data analysis. Raw fastq files were pre-processed using the pyCRAC package (Webb et al. 2014). Raw reads were demultiplexed (script pyBarcodeFilter.py) and filtered (script pyFastqDuplicateRemover.py). Processed reads were mapped to the fission yeast genome (ASM294v2.23) using the novoalign software (www.novocraft.com). Mapped files were further analysed using the pyCRAC package. Briefly, reads were mapped to the fission yeast genome annotation including all annotated coding and non-coding transcripts (www.pombase.com). Untranscribed intergenic regions were included in this genomic annotation as independent features. Taking "intergenic" features into account allows to screen

for potential binding motifs enrichment downstream of transcribed units using the pyCRAC software (see below).

Reads were mapped to genome annotation:

```
pyReadCounters.py -f myReads.novo --gtf=myAnnotation.gtf -r 0 -v --unique
```

Reads were then clustered to define regions of enriched binding:

```
pyClusterReads.py -f myReads_count_output_reads.gtf --gtf=myAnnotation.gtf --cic=10 --ch=10 --co=5 -r 0 --mutsfreq=10 -o Clusters_0_myReads_count_output_reads.gtf
```

Significant regions, or clusters, were defined with an FDR of 5%:

```
pyCalculateFDRs.py -f Clusters_0_myReads_count_output_reads.gtf --o FDR_Clusters_0_myReads_count_output_reads.gtf -c myGenome.txt -m 0.05 --min=5 --gtf=my.Annotation.gtf-
```

Motifs significantly enriched at Seb1 binding sites were extracted from significant clusters:

```
pyMotif.py -f FDR_Clusters_0_myReads_count_output_reads.gtf --gtf=myAnnotation.gtf -tab=Schizosaccharomyces_pombe.ASM294v2.23.tab --k_min=4 --k_max=8 -r 0
```

Coordinates of CRAC reads binding and of crosslinking-induced nucleotide deletions were extracted from the "myReads_count_output_reads.gtf" files (output of pyReadCounters.py) and plotted around poly(A) sites using in house R scripts as for ChIP-seq data. Statistics of read numbers binding to specific genomic features were extracted from the "myReads_hittable_reads.txt" files (output of pyReadCounters.py). Reads covering two genomic features were assigned once to each feature and therefore counted twice. Average gene analysis of CRAC data was performed as for ChIP-seq datasets using in house scripts.

SUPPLEMENTARY REFERENCES**(associated with supplementary experimental procedures, figures and tables)**

- Bähler J, Wu JQ, Longtine MS, Shah NG, McKenzie A 3rd, Steever AB, Wach A, Philippsen P, Pringle JR. Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* 14: 943-951 (1998).
- Bataille AR, Jeronimo C, Jacques PE, Laramée L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F. 2012. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* 45: 158-170.
- Beaudoin J, Laliberté J, Labbé S. 2006. Functional dissection of Ctr4 and Ctr5 amino-terminal regions reveals motifs with redundant roles in copper transport. *Microbiology* 152: 209-222.
- Coudreuse D, van Bakel H, Dewez M, Soutourina J, Parnell T, Vandenhautte J, Cairns B, Werner M, Hermand D. 2010. A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in *S. pombe*. *Curr Biol* 20: 1053-1064.
- Grenier St-Sauveur V, Soucek S, Corbett AH, Bachand F. 2013. 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* 33: 4718-4731.
- Keeney JB, Boeke JD. 1994. Efficient targeted integration at leu1-32 and ura4-294 in *Schizosaccharomyces pombe*. *Genetics* 136: 849-856.
- Larochelle M, Lemay JF, Bachand F. 2012. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic Acids Res* 40: 10240-10253.
- Lemay JF, D'Amours A, Lemieux C, Lackner DH, St-Sauveur VG, Bahler J, Bachand F. 2010. The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. *Molecular cell* 37: 34-45.

- Lemay JF, Laroche M, Marguerat S, Atkinson S, Bahler J, Bachand F. 2014. The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat Struct Mol Biol* 21: 919-926.
- Lemieux C, Marguerat S, Lafontaine J, Barbezier N, Bähler J, Bachand F. 2011. A Pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. *Mol Cell* 44: 108-119.
- Perreault A, Lemieux C, Bachand F. 2007. Regulation of the Nuclear Poly(A)-binding Protein by Arginine Methylation in Fission Yeast. *J Biol Chem* 282: 7552-7562.
- Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42: W187-91.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29: 24-26.
- Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95: 5857-5864.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7: 539.
- Sikorski RS, Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* 122: 19-27.
- Steinmetz EJ, Brow DA. 1996. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol* 16: 6993-7003.
- Steinmetz EJ, Brow DA. 2003. Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription.. *Mol Cell Biol* 23: 6339-6349.
- Thomas BJ, Rothstein R. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* 56: 619-630.

SUPPLEMENTARY TABLES

Supplementary Table S1. List of proteins identified in the Seb1-HTP purification^a

Protein IDs	MS/MS spectra count	Intensity	# of unique peptide	Protein length (aa)	% of unique sequence coverage
spQ9UTE3SEB1	1846	4,60E+11	38	620	64,4
spP31209PABP	349	2,76E+10	26	653	42,1
spP24782DBP2	426	1,14E+10	32	550	54
spQ02061RPB2	442	9,98E+09	60	1210	42,6
spO42958PAT1	266	7,55E+09	25	754	34,4
spP36594RPB1;trI7A0C5I7A0C5_SCHPM;trG9JY47G9JY47_SCHPM	347	7,53E+09	53;25;5	1752	32,7
spQ9HEQ9TCG1	84	5,87E+09	12	349	47,9
spO74835RRP5	220	4,68E+09	55	1690	35,6
spQ09330MLO3	155	3,69E+09	14	199	59,3
spP35551FBRL	119	3,56E+09	16	305	45,9
spO74191FKBP4	84	3,23E+09	12	361	36,3
spO74733CFT1	176	3,16E+09	38	1441	34,1
spQ9P7X8RPA2	253	3,07E+09	34	1174	29,8
spO13741NOP12	151	3,05E+09	15	438	35,2
spP40383XRN1	172	3,03E+09	45	1328	39
spP15398RPA1	170	3,02E+09	43	1689	32,6
spQ9UU76PUF6	143	2,85E+09	17	642	30,7
spO94268NOP2	144	2,82E+09	21	608	32,6
spO36021YEK9	152	2,70E+09	28	860	34,1
spQ9HDV4LID2	11	2,69E+09	3	1513	1,7
spO74740CFT2	137	2,46E+09	34	797	45,8
spO14007CBF5	87	2,45E+09	16	474	40,1
spO94514NOP56	101	2,26E+09	16	497	36,4
spP87115YDK9	106	2,21E+09	30	1033	35,8
spO14233RNA14	134	1,98E+09	22	733	31,8
spO94659NOG1	118	1,95E+09	25	642	41
spO13802EBP2	72	1,90E+09	13	333	40,5
spQ9P775YBEG	62	1,83E+09	12	516	24
spQ9P7G0YKY4	84	1,82E+09	17	707	25,6
spQ9HGL6BRX1	123	1,74E+09	16	295	48,1
spQ9UUG1RPF2	67	1,72E+09	14	317	49,2
spP41891GAR2	72	1,69E+09	8	500	17,4
spQ9P7S7NOP58	80	1,62E+09	17	508	45,9
spO13794YSH1	107	1,57E+09	15	757	25,1
spP40848XRN2	101	1,56E+09	25	991	30,4
spP87058LAH1	38	1,53E+09	14	298	48,3
spP09322H4	81	1,52E+09	10	103	60,2
spP78814SRP2	77	1,47E+09	9	365	28,2
spQ10202DBP3	140	1,46E+09	23	578	45,2
spQ02099RAD3	74	1,45E+09	2	2386	0,6
spO74970PRP39	99	1,44E+09	20	612	36,1
spP04913H2B1	39	1,36E+09	6	126	41,3

trS6CCW7S6CCW7_SCHPM;spQ09916HAS1	111	1,35E+09	20;20	578	40,8
spQ10222PTA1	71	1,32E+09	18	670	28,1
spQ10175FKBPH	51	1,28E+09	13	362	43,9
spO60164PESC	74	1,24E+09	16	607	26,9
spQ09191RPAB1	54	1,20E+09	8	210	33,3
spQ9UT32RL1DB	53	1,18E+09	10	373	25,5
spO74399ERB1	50	1,14E+09	17	740	34,1
spQ09820RENT1;spQ92355SEN1	74	1,11E+09	30;0	925	35,5
spO94616RPAC1	58	1,11E+09	15	348	54
spO13801YE04	58	1,07E+09	13	604	26,2
spO74393MAK5	95	1,03E+09	21	648	40,3
spO94676UTP22	86	1,03E+09	29	1097	27,4
spO74400YOCE	106	1,02E+09	21	674	32,2
spO60094DPO5	75	1,01E+09	22	959	26,4
spQ9USZ4RPA34	20	9,89E+08	10	251	39,8
spQ09775ROK1	77	9,85E+08	15	481	33,3
spO74814YJ03	55	9,30E+08	11	387	27,6
spO14086RPA49	71	9,07E+08	14	425	29,2
spQ9USZ6MRT4	32	8,80E+08	6	241	29,9
spQ9URY0YTM1	35	8,35E+08	15	440	40,5
spO42906NPA3	36	7,47E+08	8	367	23,4
spP0C8R3RS27B;spP0C016RS27A;spP0CH07RL402;spP0CH06RL401;spP0CG72UBI4P;trO74819O74819_SCHPM	49	7,08E+08	8;8;3;3;3;3	150	35,3
spO14187SPP42	82	7,01E+08	50	2363	22,1
spP40847SAP1	40	6,99E+08	7	254	31,1
spP37382RPB3	35	6,93E+08	9	297	33
spO74791GRN1	60	6,51E+08	13	470	32,1
sp O42649 SMC3_SCHPO	2	6,42E+08	2	1194	2,3
spO94692RUVB2	67	6,41E+08	18	465	41,9
spQ09903DRS1	52	6,29E+08	13	754	19,6
spO60179UTP10	52	5,97E+08	32	1649	21,8
spQ9P6P7YKCG	35	5,97E+08	8	422	25,8
spQ9UTT2RSE1	56	5,93E+08	23	1206	23,5
spO59810VGL1	85	5,82E+08	24	1291	21,1
spQ03319PRH1	69	5,81E+08	13	719	19,3
spO74862UTP23	36	5,80E+08	9	260	33,1
spQ09799KRI1	40	5,77E+08	9	598	15,2
spQ9Y7Y3YGR4	19	5,44E+08	10	233	34,3
spQ09100RNP24	32	5,41E+08	8	369	29,5
spO42947YBPC	39	5,18E+08	17	737	24,8
spQ9P3W6YLI2	35	5,13E+08	16	468	34,6
spO60055UTP20	62	5,13E+08	43	2493	17
spQ9UT24BRR2	63	5,10E+08	40	2176	20,1
spO74958MMI1	63	5,07E+08	13	488	32,4
spO94403YQFB	22	5,01E+08	7	183	35
spQ9Y824YON1	62	4,87E+08	15	1001	15,6
spQ09719DBP10	63	4,85E+08	19	848	22,8
spO42853RPAP2	27	4,77E+08	7	197	31,5
spO14307TRM61	48	4,63E+08	12	364	39,8

spO42945DHX15	72	4,55E+08	19	735	25,6
spQ9P7H0NHP2	21	4,43E+08	5	154	27,9
spO94316SN114	52	4,33E+08	20	984	24,4
spO13892YE38	47	4,32E+08	25	1076	27,1
spO94697RFC5	37	4,32E+08	10	358	33,2
spO74499LSM7	17	4,22E+08	7	113	42,5
spP13681PP11	19	4,16E+08	5	327	20,5
spQ10107MRA1	38	4,16E+08	11	359	29,8
spO94698NSA1	37	4,03E+08	12	387	41,3
spQ9P3A9YIL5	31	4,01E+08	11	773	14,6
spO14053YC47	40	3,98E+08	14	902	19,6
spO14352LSM4	14	3,94E+08	4	121	28,9
spQ9P4X3UTP7	32	3,90E+08	11	520	25,6
spO94408LSM2	19	3,87E+08	3	96	29,2
spP87177YB1C	60	3,84E+08	22	922	26,5
spQ9USN3UTP13	46	3,83E+08	20	777	29,5
spO74442UBP16	37	3,78E+08	11	457	26
spO60182RFC1	25	3,74E+08	12	934	14,8
spQ9HGL4TRM6	27	3,68E+08	12	462	27,5
spO14176PRP40	34	3,57E+08	15	695	26,2
spQ09801FIP1X	14	3,55E+08	4	344	13,4
spO42914YBI8	20	3,55E+08	8	284	31,3
spO60173DBP7	39	3,50E+08	14	709	23
spP10651H33;spP09988H31	41	3,45E+08	2;2	136	10,3
spO14011CWF8	14	3,44E+08	8	488	19,7
spO13878UTP17	33	3,41E+08	14	806	22,1
spQ09843RFC2	37	3,30E+08	12	340	30,6
spO94653BMS1	48	3,29E+08	18	1121	16,4
spO59792CIF1	39	3,29E+08	10	320	24,4
spO14232MTR4	49	3,26E+08	18	1117	17,8
spQ10295PAP	27	3,25E+08	10	566	21,2
spQ9UTN4PFS2	28	3,23E+08	8	509	17,7
spO94267SPT16	43	3,22E+08	18	1019	20,7
spP87243RIX1	34	3,05E+08	12	828	15,8
spQ9C0X6RUVB1	27	3,04E+08	12	456	32,2
spQ9HE11UTP5	47	3,04E+08	14	666	24,2
spP40231CSK2A	45	2,98E+08	8	332	24,4
spP42656RAD24	27	2,96E+08	6	270	26,3
spQ9UTE6MAK16	18	2,87E+08	7	302	22,5
spO74535CPSFX	29	2,85E+08	12	710	17,5
spO13936SPT5	31	2,80E+08	17	990	19,6
trS6BSK0S6BSK0_SCHPM;trS6BMF4S6BMF4_SCHPM;spQ09915SPT6	51	2,76E+08	18;18;18	1283	14,7
spO94449RFC4	27	2,74E+08	6	342	17
spQ9UTN3CID14	26	2,73E+08	11	684	20,2
spO74978YQL5	36	2,70E+08	6	276	22,8
spQ9C0X8RRP12	37	2,70E+08	20	1163	18,1
spP36629U2AF2	20	2,65E+08	8	517	20,5
spQ92347NOP9	43	2,65E+08	13	655	23,4

spQ9UT21YFY7	40	2,55E+08	9	366	34,7
spQ9UU79NSA2	15	2,51E+08	7	260	28,1
spQ9URY1PPK16	19	2,42E+08	3	672	9,1
spO36018BUD22	18	2,39E+08	8	388	22,4
spO74415YJL2	33	2,37E+08	25	1568	16,6
spO59678RRS1	12	2,36E+08	4	166	27,7
spQ8WZK0US107	25	2,36E+08	12	695	17,3
spO60080DBP9	30	2,29E+08	9	595	14,6
spQ9USH9YJQ1	41	2,23E+08	11	822	13
spP87123RPB11	16	2,21E+08	9	123	67,5
spQ9C1X1PWP2	49	2,21E+08	15	854	18,5
spP08096TOP2	29	2,19E+08	17	1485	13
spO94508CWF11	25	2,18E+08	19	1284	16,7
spP87312CLF1	35	2,17E+08	12	674	17,7
spQ9UT39YIQ4	19	2,09E+08	8	341	25,5
spQ9UUF2SYMC	21	2,07E+08	10	782	12,7
spO14036SMD2	27	2,06E+08	3	115	25,2
spQ9UTD7RRP15	18	2,05E+08	6	205	28,3
spQ9P6K0YLA3	13	2,03E+08	6	475	21,5
spQ9USN8YJY1	39	1,99E+08	10	1021	9,1
spO74879NOL10	21	1,95E+08	7	634	14,5
spQ9P7R9SYF1	31	1,92E+08	15	790	17,7
spO94288NOC3	24	1,89E+08	12	747	18,2
spO94536UCP12	31	1,89E+08	19	1327	15,8
spQ09804URB2	30	1,87E+08	15	1318	11,8
spO13856YEX2	21	1,80E+08	6	361	27,1
spO74777KRR1	22	1,80E+08	8	327	21,7
spQ10233RPC2	23	1,79E+08	13	1165	15,6
spO74764SPB4	23	1,73E+08	14	606	26,2
spP78963SKB1	36	1,71E+08	12	645	20
spO60161UTP4	25	1,71E+08	11	710	17,7
spO60071YBB9	33	1,71E+08	8	449	20
spQ10299CLP1	11	1,68E+08	5	456	13,8
spQ76PD3DBP6	38	1,65E+08	10	604	17,5
spQ9Y7U7SND1	28	1,65E+08	16	878	19,9
spQ09870RCL1	24	1,65E+08	8	363	19
spO74968RU1A	16	1,61E+08	10	249	44,6
spO14235POF5	4	1,60E+08	2	348	14,9
spO13759CSX1	3	1,56E+08	2	632	4,4
spQ9Y8G3MEX67	32	1,56E+08	10	596	21,1
spP78750UTP18	19	1,49E+08	6	519	11,2
spQ10237YD14	11	1,49E+08	6	638	10,5
spO60114YG65	33	1,48E+08	14	1428	10,7
spO14003RFC3	19	1,48E+08	7	342	22,2
spO13823IMP4	7	1,46E+08	5	289	21,1
spQ9URY7FMDA	3	1,45E+08	2	410	18,5
spP87144SYTC	24	1,42E+08	6	703	9
spQ1MTQ9NIP7	22	1,41E+08	4	180	24,4
spO13971SGD1	17	1,40E+08	11	775	17,3

spQ9HFE6YNW6	16	1,39E+08	9	478	20,3
spO13877RPAB5	15	1,38E+08	2	71	29,6
spQ9USX8RU2A	24	1,38E+08	6	239	26,8
spQ9USU7YHHB	5	1,38E+08	4	310	19,4
spO74454STS5	31	1,34E+08	14	1066	14,1
spQ12381PRP1	22	1,34E+08	5	906	5,3
spQ9P7P1RPA14	6	1,31E+08	3	147	19,7
spQ9P6K7YLF4	19	1,27E+08	8	450	26
spQ96WV9CDK9	25	1,26E+08	8	591	14,2
spO94365UTP15	13	1,26E+08	6	494	24,3
spP87137YDM6	21	1,22E+08	8	929	9,8
spO14236NOG2	23	1,20E+08	8	537	15,8
spO42964IWS1	23	1,19E+08	11	428	32,9
spO14326PUB3	21	1,19E+08	6	786	8,1
spO13845RSD1	20	1,16E+08	8	603	16,9
spQ09867EFG1P	15	1,16E+08	7	183	34,4
spQ9US25HRP1	29	1,16E+08	13	1373	9,7
spO13874YE19	16	1,15E+08	6	528	13,3
spO14206YDD3	14	1,15E+08	8	389	20,3
spO13910MPP10	18	1,08E+08	6	598	10,5
spO60072MUG81	13	1,08E+08	15	1935	9,3
spO13775SYEC	18	1,06E+08	10	716	12,7
spO42832SPB1	25	1,05E+08	7	802	10,5
spQ9UTP9DBP4	18	1,03E+08	6	735	8,3
spQ9Y7Z1NOP16	6	1,02E+08	4	209	23,4
spQ9USV4MEP33	9	9,87E+07	9	292	33,6
spQ9P7C0YKU5	19	9,77E+07	5	524	10,5
spO43071PRP17	14	9,74E+07	6	558	21,3
spQ10146RRP6	13	9,62E+07	11	777	17,1
spO94432YHKF	9	9,54E+07	3	474	7,8
spO74555BBP	7	9,47E+07	2	587	6
spO13651SYIC	20	9,38E+07	12	1064	13,8
spO13935TRM4B	11	9,37E+07	7	685	11,8
spO74340DCA13	16	9,33E+07	5	436	13,1
sp O13765 DBR1_SCHPO	2	8,92E+07	2	478	4,4
spO74356YH16	8	8,81E+07	6	585	10,9
spQ9UUI1LSM6	8	8,70E+07	4	75	54,7
spO94666RPC1	26	8,63E+07	13	1405	10,7
spO14459RPB7	12	8,60E+07	2	172	15,1
spQ10272IP13	11	8,52E+07	6	446	15,2
spP42657RAD25	7	8,48E+07	3	270	11,5
spQ10362SET3	4	8,45E+07	2	859	2,3
spQ9UTD1YTH1	7	8,34E+07	3	170	15,3
spQ9P7S3SSR3	9	8,31E+07	6	425	12,5
trQ8WZJ6Q8WZJ6_SCHPM	11	8,06E+07	5	124	16,9
spQ9P6R9CWC22	20	8,01E+07	6	887	8,3
spQ09181DHH1	17	8,01E+07	7	485	17,3
spQ09826SDS23	14	7,91E+07	5	408	19,6
spQ09729YA4C	11	7,85E+07	8	596	17,1

spO14268YFPD	12	7,84E+07	6	312	18,6
spQ9UUC6SMD3	10	7,81E+07	3	97	24,7
spQ9Y7M6RRP1	11	7,68E+07	4	217	15,7
spO60059YG58	14	7,60E+07	7	661	9,5
spQ09868CIP2	11	7,60E+07	6	576	12,2
spP27638SPK1	16	7,47E+07	6	372	19,6
spP36616DSK1	23	7,47E+07	8	544	19,5
spQ10414MU161;trP78893P78893_SCHPM	9	7,45E+07	6;1	561	11,4
spO74894PRS6B;spO60058AFG2	9	7,44E+07	6;0	389	26,7
spO13615PRP46	9	7,41E+07	6	473	16,3
spQ9UUI4YIY9	8	7,40E+07	4	419	11,5
sp Q9UTJ8 RAD50_SCHPO	4	7,36E+07	4	1285	4
spO13919PST2	11	7,34E+07	9	1075	9,4
spO14148INO80	18	7,34E+07	13	1604	9,1
spO13799YE02	15	7,33E+07	3	1030	3,3
spP39964CEF1	14	7,15E+07	7	757	9,2
spO14253NCBP1	17	7,03E+07	9	780	11,9
spO14180YDS4	7	6,95E+07	3	306	9,2
spP37202DIS3	14	6,92E+07	11	970	14,2
spQ9UUI3SA145	13	6,87E+07	5	601	8,7
spO94620CWF17	8	6,66E+07	6	340	26,8
trS6BSJ4S6BSJ4_SCHPM;spQ9P7T8YIW2	9	6,60E+07	9;9	1717	5,2
spO13620MRD1	14	6,59E+07	10	833	16,4
spO42936LAS1	15	6,57E+07	4	470	11,7
spQ9UTA4POPI	15	6,47E+07	6	698	8,2
spQ10163RSMB	10	6,45E+07	3	147	21,8
spO94260G3BP	7	6,36E+07	5	434	14,5
spQ9USY2YOW4	16	6,26E+07	7	1014	8,9
spO14021PRW1	7	6,20E+07	4	431	11,4
spQ9P3A7CDC48	13	6,18E+07	7	815	12,1
spO74828ESF1	10	6,04E+07	4	682	8,2
spO60188UTP6	10	6,03E+07	4	488	9,6
spO13824NMD2	10	5,73E+07	10	1049	11,2
spP41838RAE1	5	5,72E+07	5	352	21
spQ9HDX7YK1D	6	5,70E+07	6	529	17
spQ10353RLP24	9	5,60E+07	5	192	25,5
spP23880PP12;spP78968PPZ	7	5,50E+07	3;1	322	7,8
spO42868SSU72	9	5,48E+07	3	197	16,2
spO59706SAP61	13	5,47E+07	7	492	12,6
spQ12126CRK1	3	5,35E+07	3	335	15,2
spO13956TSR1	14	5,24E+07	6	783	8,6
spQ09699SNF5	4	5,22E+07	2	632	5,9
spP40232CSK2B	8	5,17E+07	3	231	18,2
sp P87112 NOT1_SCHPO	9	5,13E+07	10	2100	5,7
spQ10277MSA1	6	5,11E+07	2	533	4,5
spQ9US15YK62	3	5,10E+07	4	231	14,3
spG2TRS0YH0H	7	5,10E+07	4	81	19,8
spO74855NLE1	11	4,91E+07	5	502	13,5
spQ10257RRP8	7	4,88E+07	3	318	11

spO94529POB3	5	4,87E+07	6	512	13,9
spP87173LSM1	7	4,79E+07	4	140	27,1
spO43036RPA43	3	4,79E+07	2	173	9,2
spO14077MU138	11	4,71E+07	2	969	2,8
spQ9Y7R7RL1DA	18	4,52E+07	2	284	6,7
spQ1MTS0YBH9	8	4,51E+07	5	597	10,7
spQ92462PUB1	8	4,51E+07	3	767	4
spQ10340YBLJ	11	4,44E+07	6	515	11,5
spO60180ARX1	6	4,43E+07	5	417	15,3
spQ9Y7N1YCK3	6	4,41E+07	5	240	19,2
spO94567SYNC	13	4,20E+07	5	568	7,4
spO13953EAF3	8	4,14E+07	4	337	12,2
spQ92399RPAB3	8	4,06E+07	2	125	16,8
spQ9Y7M4LSM3	4	4,04E+07	2	93	21,5
spO43040CTF1	5	3,99E+07	2	363	5,5
spQ10106UTP11	7	3,95E+07	5	249	24,1
spQ9US05BFR2	8	3,86E+07	3	452	7,1
spQ9P7H1FLP1	10	3,84E+07	5	537	11,2
spQ9Y7N0BDF1	5	3,81E+07	2	578	7,3
spQ9P7C7PRP11	8	3,73E+07	4	1014	4,1
spQ9P3U1YKX5	6	3,72E+07	5	464	15,3
spQ9USM4LUC7	12	3,58E+07	5	264	15,2
spQ10342SDA1	6	3,55E+07	4	719	9
spQ9P7F2CAO1	8	3,53E+07	5	712	5,6
sp Q09779 THO2_SCHPO	6	3,52E+07	6	1628	5,7
spO74918RRP45	3	3,51E+07	3	291	13,1
spO94700MBF1	5	3,51E+07	3	148	25
sp O94387 YGSA_SCHPO	7	3,50E+07	3	1944	2,1
spO94396PUS1	10	3,45E+07	4	534	7,1
spO14030YEM8	5	3,44E+07	2	689	3
spQ09176U2AF1	5	3,42E+07	3	216	12,5
spG2TRQ9TAM10	3	3,41E+07	2	168	14,9
spO42943YBP8	10	3,40E+07	6	618	12,6
sp P0CT33 TLH1_SCHPO;sp Q1RKN3 TLH2_SCHPO	7	3,37E+07	6;6	1887	3,1
sp P87154 DPOE_SCHPO	9	3,36E+07	6	2199	3,7
spO74919RNC1	9	3,35E+07	4	398	11,8
spO74465HRR1	7	3,28E+07	5	999	5,7
spO60152UBR1	10	3,26E+07	8	1958	4,6
spQ09731BU107	7	3,24E+07	7	962	8
spQ9P7M5NOP10	5	3,20E+07	2	64	40,6
spO94714SPT2	3	3,17E+07	3	406	9,6
spQ09817NMD3	5	3,09E+07	4	498	9,8
spO94621US109	3	3,07E+07	2	365	7,1
spQ9USV3SPF27	5	3,07E+07	4	187	21,4
spO74453SKB15	7	2,95E+07	2	341	7,6
spQ9UU96GRC3	10	2,93E+07	5	736	7,6
spQ10055FAL1	7	2,87E+07	7	394	18,5
spO13791SLT1	5	2,84E+07	4	433	10,9
spO74445PRS10	6	2,82E+07	2	388	6,4

spO14044PNO1	4	2,79E+07	3	241	13,7
spP78794CWC15	3	2,78E+07	3	265	12,1
spQ09818YAC4	4	2,69E+07	5	489	14,7
spQ9P783RRB1	5	2,60E+07	5	480	11,5
spQ10204RRN9	5	2,50E+07	4	250	16
spP04551CDK1;spO60145PPK23	8	2,46E+07	2;0	297	5,4
spO14013RRN5	6	2,45E+07	3	556	5,6
sp P87315 HIR3_SCHPO	3	2,42E+07	3	1630	1,8
spO13829RU17	5	2,41E+07	4	261	18,8
spQ9Y805CWC25	8	2,40E+07	5	376	16,8
sp Q09750 PST1_SCHPO	2	2,39E+07	2	1522	3,5
spO43051NOP14	5	2,39E+07	5	827	7
spQ10267UPF3	5	2,38E+07	2	278	7,6
spO13862RDP1	4	2,31E+07	3	478	9,2
sp O14014 RGA3_SCHPO	4	2,30E+07	3	969	8,5
sp O74755 PST3_SCHPO	2	2,29E+07	2	1154	2,6
spO94569YGDC	4	2,24E+07	2	594	6,6
spQ96WW3YNH5	3	2,20E+07	2	117	15,4
spO59702CLR6	7	2,19E+07	2	405	4,2
spQ09882PRP45	3	2,15E+07	2	557	3,4
spO14098CTK1	9	2,13E+07	4	593	5,6
spO43006MU178	5	2,13E+07	3	225	13,8
spO74825RPB4	7	2,09E+07	4	135	44,4
spP17871TBP	4	2,05E+07	3	231	12,1
spQ10099SEH1	5	2,04E+07	2	339	6,8
spP36612PRS4	5	2,00E+07	3	448	7,4
sp P87311 MUB1_SCHPO	2	1,99E+07	2	574	8,4
spQ9P7W8RSC9	6	1,94E+07	5	780	6,4
sp Q09872 YAG9_SCHPO	4	1,91E+07	3	977	3,3
spP0CAN8NXT1	3	1,88E+07	3	115	21,7
spQ9UT50YFR3	4	1,87E+07	3	485	7,4
spO13959RBX1	4	1,78E+07	2	107	15,9
sp Q10059 CDC12_SCHPO	5	1,78E+07	5	1841	2,8
spQ10313TRNL	10	1,77E+07	5	787	6,4
spO94372YG06	4	1,76E+07	3	485	3,7
sp O14139 HRP3_SCHPO	6	1,72E+07	4	1388	3
spQ76PD2YJ01	5	1,71E+07	2	732	2,7
spO74627CG1C	2	1,71E+07	2	342	7,6
sp O74360 RGA4_SCHPO	4	1,69E+07	5	933	7
spQ10437BUN62	4	1,67E+07	3	543	5,7
spO59719YBHB	5	1,67E+07	2	868	2,6
spO14026SET2	5	1,66E+07	2	798	2,1
spO42923CIP1	4	1,66E+07	3	490	6,1
spP87126CWC2	4	1,65E+07	4	388	12,1
spO13992MUG56	6	1,59E+07	3	965	5,4
sp Q9UTN6 SNF21_SCHPO	5	1,57E+07	4	1199	3,3
spQ9Y800MUG97	6	1,57E+07	3	345	7,5
sp O94600 TTI1_SCHPO	2	1,56E+07	2	1098	4,7
spQ9C106RPC3	4	1,56E+07	4	591	6,9

spO59821CSL4	3	1,51E+07	3	181	18,2
spQ10193SRP1	5	1,51E+07	3	275	9,8
spO14338YB33	5	1,50E+07	4	891	4,2
spO13836DXO	7	1,43E+07	4	352	11,9
spO13788SSR1	4	1,43E+07	3	527	6,1
spO42643PRP22	6	1,42E+07	2	1168	2,1
sp P41389 MCM5_SCHPO	3	1,42E+07	2	720	3,1
spO14040DI3L2	6	1,41E+07	4	927	4,2
spO42648YF98	2	1,38E+07	3	341	11,7
spP41836PRS8;spQ9P7Q4SEC18	3	1,37E+07	2;1	403	6,2
spO42931PRS7	3	1,31E+07	2	438	3,9
spQ09177RPAC2	5	1,30E+07	3	125	24,8
spQ9USU2DIM1	4	1,27E+07	2	307	13,4
spQ92368RCD1	3	1,26E+07	3	283	10,6
sp O74424 NU211_SCHPO	2	1,26E+07	2	1837	1,3
spO43082RR14C	7	1,26E+07	3	233	8,2
spQ9HGM5DBP8	4	1,25E+07	3	453	8,2
spQ9USJ8CTK3	4	1,25E+07	2	218	21,1
spP31411VATB	3	1,21E+07	2	503	5,2
spQ09819YAC5	5	1,19E+07	2	404	5,9
sp P33277 GAP1_SCHPO	3	1,17E+07	2	766	5,6
spQ9UTA5TYW3	2	1,17E+07	3	237	11,8
spQ9Y820YON7	5	1,14E+07	2	442	5,2
sp Q01663 AP1_SCHPO	2	1,07E+07	2	552	4,3
sp Q9HDY4 YK16_SCHPO	5	1,06E+07	3	1183	2,5
spQ9UUE6SYKC	5	1,06E+07	2	190	4,1
sp O13798 CID16_SCHPO	3	1,05E+07	3	1202	2,7
spQ10110RRN3	3	1,02E+07	2	599	10,5
spO74370ISY1	3	9,96E+06	2	217	8,8
sp O74431 ATC9_SCHPO	4	9,90E+06	2	1315	3,8
sp P41003 SMC2_SCHPO	5	9,89E+06	4	1172	3,3
sp P41004 SMC4_SCHPO	3	9,80E+06	3	1324	3,1
sp Q09833 YAD7_SCHPO	2	9,78E+06	2	527	3,6
spO74857RPC4	2	9,77E+06	2	330	5,8
spO59811REI1	6	9,26E+06	2	463	4,8
spQ9C0Z3SLX9	4	8,99E+06	2	172	12,2
spP40235HHP1	3	8,92E+06	2	365	7,1
sp Q09733 GEF2_SCHPO	2	8,80E+06	2	1101	3,2
spO59670PM14	2	8,74E+06	2	115	14,8
sp O14325 YB7A_SCHPO	3	8,66E+06	3	779	4,1
spQ9P7S9IES4	2	8,53E+06	2	194	10,3
trQ152U6Q152U6_SCHPM;spP22192PAC1	4	8,39E+06	2;2	363	5,8
spO14278MU129	2	8,32E+06	2	302	8,9
trK7PD52K7PD52_SCHPM;spG2TRN4ERH	2	8,31E+06	2;2	104	15,4
spO94400YQF7	3	8,29E+06	3	571	4,4
spP30316DPOD	5	8,20E+06	3	1086	5,2
sp Q1MTM9 YFN1_SCHPO	4	8,14E+06	3	977	9,5
spQ10183YAW8	3	7,85E+06	2	259	10,4
spQ9UTB8INI1	5	7,82E+06	2	117	14,5

spQ09873WDR21	3	7,79E+06	2	420	4
spO74741DIOXL	5	7,74E+06	3	297	16,5
spQ9USZ2YNR7	6	7,41E+06	2	955	3,2
sp Q9UUA2-2 PIF1_SCHPO;sp Q9UUA2 PIF1_SCHPO	3	7,31E+06	3;3	785	4,6
spO74883RPC5	2	7,20E+06	2	242	7,9
spQ10179SYG	4	7,01E+06	2	652	4,9
spO75005SYV	5	6,93E+06	3	980	3,7
spO74880MCES	2	6,87E+06	2	360	5,8
sp Q9Y7V2 BYE1_SCHPO	2	6,76E+06	2	721	3,1
sp Q9UTH9 DUS3_SCHPO	2	6,71E+06	2	617	7
spO59679GEF3	3	6,66E+06	2	525	2,3
spO13815MUG52	3	6,07E+06	2	105	37,1
sp O43065 MOT1_SCHPO	2	6,07E+06	2	1953	1
sp O94374 IMA2_SCHPO	2	5,91E+06	2	539	3,7
sp O60155 PRS1_SCHPO	3	5,89E+06	2	716	2,9
spO74522ADN3	5	5,58E+06	3	964	4,3
sp O74349 LTN1_SCHPO	4	5,53E+06	4	1610	3,7
spP87120RNP3	4	5,52E+06	2	235	8,5
sp Q09909 CWF19_SCHPO	2	5,13E+06	2	639	3,1
sp P87314 HIR1_SCHPO	3	5,00E+06	2	932	2,4
sp Q9URT2 YJU3_SCHPO	3	4,82E+06	2	752	2,3
sp O74485 CUE3_SCHPO	2	4,71E+06	2	581	3,4
spQ9USL1YJKA	2	4,60E+06	2	309	6,5
spO94718IPI1	2	4,46E+06	2	413	5,3
sp O74985 NAT1_SCHPO	3	4,39E+06	3	729	4
spP87310MED10	3	4,33E+06	2	144	11,8
sp O74363 YND2_SCHPO	2	4,27E+06	2	695	3
sp Q9Y7K2 TOR2_SCHPO	2	4,15E+06	2	2337	0,9
sp Q09822 CDC15_SCHPO	3	3,80E+06	2	927	2,8
spQ09888MU160	2	3,21E+06	2	434	6,7
spQ09728CUF1	3	3,20E+06	2	411	3,9
sp O74957 AGO1_SCHPO	3	3,20E+06	2	834	4,2
sp Q09778 TSC1_SCHPO	2	3,09E+06	2	899	4,1
sp O14114 YEJ1_SCHPO;sp Q9C0W2 YHI5_SCHPO	2	3,05E+06	2;1	1190	1,8
spQ9UTM0IEC1	2	2,96E+06	2	249	10,4
spO94418MUG87	5	2,92E+06	2	851	2,4
spP41412RES2	5	2,87E+06	2	657	3,3
sp O59712 YBH4_SCHPO	2	2,83E+06	2	867	5
spG2TRS9YBJ2;spQ9HDT8YJ51;spQ9Y7S0YQO1	2	2,13E+06	2;1;1	305	8,9
sp Q92372 RFA1_SCHPO	2	2,12E+06	2	609	2,6
sp O94527 IQW1_SCHPO	2	1,32E+06	2	809	2,3

^a Exclusion: Ribosomal proteins, proteasome components, heat shock proteins/chaperones, cytoskeletal proteins, cell wall proteins, protein trafficking factors, metabolic enzymes, translation proteins, mitochondrial proteins, vacuolar proteins, peroxisomal proteins, protein with less than 2 spectral counts, protein with less than 2 unique peptides.

Supplementary Table S2. List of Seb1-HTP -associated proteins with putative functions in 3'-end processing and transcription termination of mRNAs.

<i>S. cerevisiae</i> putative orthologue	Gene name	Systematic	MS/MS spectra count	Intensity	# of unique peptide	Protein length (aa)	% of unique sequence coverage
Gene ontology : mRNA cleavage factor complex							
Cft1	<i>cft1</i>	<i>SPBC1709.08</i>	176	3,16E+09	38	1441	34,1
Cft2	<i>cft2</i>	<i>SPBC1709.15c</i>	137	2,46E+09	34	797	45,8
Rna14	<i>rna14</i>	<i>SPAC6F12.17</i>	134	1,98E+09	22	733	31,8
Brr5/Syc1	<i>ysh1</i>	<i>SPAC17G6.16c</i>	107	1,57E+09	15	757	25,1
Pta1	<i>pta1</i>	<i>SPAC1071.01c</i>	71	1,32E+09	18	670	28,1
Rna15	<i>SPAC644.16</i>	<i>SPAC644.16</i>	35	5,97E+08	8	422	25,8
Glc7	<i>dis2</i>	<i>SPBC776.02c</i>	19	4,16E+08	5	327	20,5
Fip1	<i>iss1</i>	<i>SPAC22G7.10</i>	14	3,55E+08	4	344	13,4
Pla1	<i>pla1</i>	<i>SPBC646.04</i>	27	3,25E+08	10	566	21,2
Pfs2	<i>pfs2</i>	<i>SPAC12G12.14c</i>	28	3,23E+08	8	509	17,7
N/A	<i>ppn1</i>	<i>SPCC74.02c</i>	29	2,85E+08	12	710	17,5
Swd2	<i>swd22</i>	<i>SPAC824.04</i>	19	2,09E+08	8	341	25,5
Clp1	<i>SPAC22H10.05c</i>	<i>SPAC22H10.05c</i>	11	1,68E+08	5	456	13,8
Pcf11	<i>pcf11</i>	<i>SPAC4G9.04c</i>	11	1,49E+08	6	638	10,5
Hrp1	<i>mis2/nab4</i>	<i>SPBC660.15</i>	9	9,54E+07	3	474	7,8
Yth1	<i>yth1</i>	<i>SPAC227.08c</i>	7	8,34E+07	3	170	15,3
Ssu72	<i>ssu72</i>	<i>SPAC3G9.04</i>	9	5,48E+07	3	197	16,2
Rna15	<i>ctf1</i>	<i>SPBC3B9.11c</i>	5	3,99E+07	2	363	5,5
Mpe1	<i>mpe1</i>	<i>SPBP8B7.15c</i>				No data	
Gene ontology : termination of RNAPII transcription, poly(A)-coupled							
Rat1	<i>dhp1</i>	<i>SPAC26A3.12c</i>	101	1,56E+09	25	991	30,4
Rtt103	<i>rhn1</i>	<i>SPBC337.03</i>	55	9,30E+08	11	387	27,6
Rai1	<i>din1</i>	<i>SPAC19D5.06c</i>	7	1,43E+07	4	352	11,9

Supplementary Table S3. List of the top 100 genes displaying 3' UTR lengthening upon Seb1-depletion.

3'READS analyses (WT) and seb1 mutant (P_{nmt1}-seb1) cells grown in the presence of thiamine^a											
Gene name	Systematic name	Gene type	Chromosome ID	Strand of transcription	Proximal p(A) isoform position	Distal p(A) isoform position	Proximal p(A) isoform in WT (RPM)^b	Proximal p(A) isoform in seb1 mutant (RPM)	Distal p(A) isoform in WT (RPM)	Distal p(A) isoform in seb1 mutant (RPM)	SS^b
<i>agn1</i>	<i>SPAC14C4.09</i>	mRNA	I	+	5245904	5246165	2733,7	407,9	1,1	523,9	999
<i>arx1</i>	<i>SPBC23E6.05</i>	mRNA	II	+	3851546	3851578	5302,5	3145,1	564,6	2317,9	999
<i>atp1</i>	<i>SPAC14C4.14</i>	mRNA	I	+	5258922	5259527	663,3	359,1	21,0	4717,3	999
<i>but2</i>	<i>SPBC3D6.02</i>	mRNA	II	+	1271462	1271729	1659,1	167,3	3,3	1568,5	999
<i>cdc48</i>	<i>SPAC1565.08</i>	mRNA	I	+	1309148	1309314	409,1	532,0	40,8	5103,3	999
<i>erg11</i>	<i>SPAC13A11.02c</i>	mRNA	I	-	574983	574641	667,7	28,8	97,0	1247,6	999
<i>fba1</i>	<i>SPBC19C2.07</i>	mRNA	II	+	1689546	1689579	6150,0	2109,9	7484,3	6971,3	999
<i>gly1</i>	<i>SPAC23H3.09c</i>	mRNA	I	-	2507755	2507707	14391,8	546,4	991,9	630,4	999
<i>lys3</i>	<i>SPAC227.18</i>	mRNA	I	+	531723	531874	1646,4	371,6	243,7	1418,1	999
<i>met26</i>	<i>SPAC9.09</i>	mRNA	I	+	1481172	1481221	19285,7	2901,3	823,7	3205,9	999
<i>pabp</i>	<i>SPAC57A7.04c</i>	mRNA	I	+	1545065	1545202	401,9	52,0	662,7	3046,1	999
<i>pfk1</i>	<i>SPBC16H5.02</i>	mRNA	II	-	2297233	2297087	561,3	145,4	0,6	555,8	999
<i>pgi1</i>	<i>SPBC1604.05</i>	mRNA	II	-	3921724	3921572	1592,9	233,1	46,9	487,5	999
<i>pgk1</i>	<i>SPBC14F5.04c</i>	mRNA	II	-	4157821	4157750	1522,3	1106,0	57,9	5070,7	999
<i>plb1</i>	<i>SPAC1A6.04c</i>	mRNA	I	-	1074334	1074309	14289,2	11888,6	1279,2	8522,3	999
<i>ppi1</i>	<i>SPBC28F2.03</i>	mRNA	II	+	1576915	1577153	3707,9	945,6	9,9	4580,7	999
<i>prl53</i>	<i>SPNCRNA.53</i>	ncRNA	I	-	4009330	4008054	1788,1	300,2	33,1	2360,5	999
<i>rpl1001</i>	<i>SPBC18E5.04</i>	mRNA	II	+	2082212	2082309	1838,3	85,8	72,2	595,9	999
<i>rpl1101</i>	<i>SPAC26A3.07c</i>	mRNA	I	-	3346132	3345897	1272,0	493,2	2,2	904,2	999
<i>rpl1202</i>	<i>SPCC31H12.04c</i>	mRNA	III	-	627062	626948	11071,5	3594,4	175,9	1454,4	999
<i>rpl14</i>	<i>SPAC1805.13</i>	mRNA	I	+	2795647	2795851	3875,6	1079,1	0,6	518,9	999
<i>rpl1602</i>	<i>SPAC23A1.11</i>	mRNA	I	+	4098019	4098177	1245,0	37,0	2,8	514,5	999
<i>rpl1702</i>	<i>SPCC364.03</i>	mRNA	III	-	482155	481882	912,5	109,7	8,3	631,7	999
<i>rpl1802</i>	<i>SPAPB17E12.13</i>	mRNA	I	+	1286325	1286461	6076,6	1539,0	2,2	575,9	999
<i>rpl1901</i>	<i>SPBC56F2.02</i>	mRNA	II	-	4123115	4122971	2649,9	829,7	2,8	1265,2	999
<i>rpl22</i>	<i>SPAC11E3.15</i>	mRNA	I	+	5315415	5315581	381,0	184,2	17,1	8934,6	999
<i>rpl2402</i>	<i>SPCC330.14c</i>	mRNA	III	-	138059	137898	2057,7	626,0	0,6	408,6	999
<i>rpl2701</i>	<i>SPBC685.07c</i>	mRNA	II	-	2780156	2779870	2710,5	547,1	3,3	444,3	999

<i>rpl301</i>	<i>SPAC17A5.03</i>	mRNA	I	+	1757327	1757394	114202,0	97967,8	13,8	5288,8	999
<i>rpl31</i>	<i>SPAC890.08</i>	mRNA	I	+	5018621	5018669	12969,8	4712,9	757,0	1707,0	999
<i>rpl3401</i>	<i>SPAC23A1.08c</i>	mRNA	I	-	4093344	4093253	1789,7	321,5	9,4	649,8	999
<i>rpl3802</i>	<i>SPAC30D11.12</i>	mRNA	I	-	1096859	1096711	2932,7	942,5	6,1	594,1	999
<i>rpl39</i>	<i>SPCC663.04</i>	mRNA	III	+	1641337	1641463	257,5	18,2	74,4	1722,6	999
<i>rpl701</i>	<i>SPBC18H10.12c</i>	mRNA	II	-	1790970	1790871	1175,0	891,1	2,2	631,0	999
<i>rpl8</i>	<i>SPBC29A3.04</i>	mRNA	II	+	2046442	2046591	2948,7	2063,5	16,0	1966,4	999
<i>rpl801</i>	<i>SPAC1F7.13c</i>	mRNA	I	-	4250620	4250430	12002,2	1446,3	49,6	20020,4	999
<i>rpp0</i>	<i>SPCC18.14c</i>	mRNA	III	-	1980570	1980482	5720,4	1610,5	16,5	726,9	999
<i>rpp101</i>	<i>SPAC644.15</i>	mRNA	I	+	2700989	2701239	449,9	26,3	25,9	1378,0	999
<i>rps002</i>	<i>SPAPJ698.02c</i>	mRNA	I	-	4036639	4036531	1584,6	100,3	156,0	493,8	999
<i>rps1402</i>	<i>SPBC18H10.13</i>	mRNA	II	+	1792936	1793002	1213,0	222,5	34,2	1285,9	999
<i>rps1501</i>	<i>SPCC1393.03</i>	mRNA	III	+	798856	798947	741,6	85,2	13,2	1244,5	999
<i>rps1601</i>	<i>SPBC18H10.14</i>	mRNA	II	+	1794699	1794758	11858,3	2232,1	254,2	1253,3	999
<i>rps1701</i>	<i>SPBC839.05c</i>	mRNA	II	-	603674	603344	5002,0	1795,9	1,7	641,0	999
<i>rps21</i>	<i>SPBC18E5.06</i>	mRNA	II	+	2084250	2084323	2805,9	614,7	77,7	4344,5	999
<i>rps2401</i>	<i>SPAC17G6.06</i>	mRNA	I	+	3599647	3599786	2968,6	1093,5	3907,5	6923,7	999
<i>rps2801</i>	<i>SPAC25G10.06</i>	mRNA	I	+	4306391	4306518	5177,3	527,0	53,5	6195,6	999
<i>rps3</i>	<i>SPBC16G5.14c</i>	mRNA	II	-	4238266	4238157	535,4	258,2	1,7	795,8	999
<i>rps7</i>	<i>SPAC18G6.14c</i>	mRNA	I	-	2241853	2241746	5533,5	1136,1	119,1	1881,8	999
<i>rps802</i>	<i>SPAC521.05</i>	mRNA	I	+	850160	850434	1516,8	201,2	5,0	773,3	999
<i>sks2</i>	<i>SPBC1709.05</i>	mRNA	II	+	1108405	1108506	3373,3	436,8	1252,2	3904,6	999
<i>sod1</i>	<i>SPAC821.10c</i>	mRNA	I	-	1002890	1002827	945,0	80,8	111,9	568,4	999
<i>SPAC17C9.12</i>	<i>SPAC17C9.12</i>	mRNA	I	-	4478306	4477847	792,3	237,5	2,2	431,1	999
<i>SPAC24C9.12c</i>	<i>SPAC24C9.12c</i>	mRNA	I	-	3069320	3069171	30699,0	7546,0	46,3	1118,5	999
<i>SPAC6F12.04</i>	<i>SPAC6F12.04</i>	mRNA	I	+	1316644	1316689	1005,7	33,2	2460,2	2363,0	999
<i>SPAC806.11</i>	<i>SPAC806.11</i>	mRNA	I	-	237049	236880	898,7	92,7	6,1	1406,8	999
<i>SPAC922.04</i>	<i>SPAC922.04</i>	mRNA	I	+	5480405	5480489	748,8	20,1	65,6	315,8	999
<i>SPBC2G5.05</i>	<i>SPBC2G5.05</i>	mRNA	II	+	2588631	2588978	1498,6	29,5	6,6	1142,4	999
<i>SPCC1281.06c</i>	<i>SPCC1281.06c</i>	mRNA	III	-	1392558	1392482	1442,9	1295,3	1,1	725,6	999
<i>swi3</i>	<i>SPBC30D10.04</i>	mRNA	II	-	3090891	3090846	1150,2	322,1	372,2	1074,7	999
<i>tif51</i>	<i>SPAC26H5.10c</i>	mRNA	I	-	4141286	4141203	876,7	1826,6	40,2	3836,3	999
<i>ubi4</i>	<i>SPBC337.08c</i>	mRNA	II	-	1044309	1044239	784,0	594,1	658,3	3526,1	999
<i>wos2</i>	<i>SPAC9E9.13</i>	mRNA	I	-	4468188	4467785	277,9	14,4	16,5	3933,4	999
<i>zip1</i>	<i>SPAC25G10.03</i>	mRNA	I	+	4301296	4301339	3737,2	2238,3	737,2	3175,8	999
<i>rpl3201</i>	<i>SPBC16C6.11</i>	mRNA	II	+	4352334	4352405	178,1	49,5	39,1	1849,8	299
<i>hht3</i>	<i>SPBC1105.11c</i>	mRNA	II	-	3528852	3528764	453,8	18,8	1122,6	2480,2	297
<i>ipp1</i>	<i>SPAC23C11.05</i>	mRNA	I	+	2141448	2141525	11719,9	747,0	114,1	227,5	288
<i>tim11</i>	<i>SPBC106.05c</i>	mRNA	II	-	383375	383296	3686,4	1357,9	297,7	774,5	282
<i>SPAC1006.07</i>	<i>SPAC1006.07</i>	mRNA	I	+	5087016	5087210	1132,5	79,6	635,7	775,2	282
<i>fis1</i>	<i>SPBC11G11.01</i>	mRNA	II	+	1759286	1759417	231,0	21,3	151,6	1442,5	277
<i>rpp202</i>	<i>SPBC23G7.15c</i>	mRNA	II	-	2127992	2127861	2295,3	504,4	1,7	230,6	258
<i>sup35</i>	<i>SPCC584.04</i>	mRNA	III	+	1537598	1537930	1441,8	21,9	259,7	265,7	245
<i>cdc42</i>	<i>SPAC110.03</i>	mRNA	I	+	1917529	1917675	521,6	32,0	270,2	621,6	238
<i>SPCC1620.06c</i>	<i>SPCC1620.06c</i>	mRNA	III	-	2153635	2152820	489,1	290,8	19,8	592,2	236
<i>rds1</i>	<i>SPAC343.12</i>	mRNA	I	+	1669646	1669788	622,5	3841,9	4,4	2778,5	233
<i>rpl2802</i>	<i>SPCC5E4.07</i>	mRNA	III	+	657105	657265	4959,0	1146,1	0,6	198,0	232
<i>SPCC1827.03c</i>	<i>SPCC1827.03c</i>	mRNA	III	-	2378387	2378252	135,6	7,5	103,1	1675,0	225
<i>cdc8</i>	<i>SPAC27F1.02c</i>	mRNA	I	-	4319246	4319050	205,7	43,2	56,2	864,8	224

<i>rad24</i>	<i>SPAC8E11.02c</i>	mRNA	I	+	3385576	3385737	613,1	82,7	1,7	194,3	214
<i>gdh1</i>	<i>SPCC622.12c</i>	mRNA	III	-	1419288	1419130	114,7	32,6	22,1	1387,4	209
<i>sla1</i>	<i>SPAC57A10.10c</i>	mRNA	I	-	1382424	1382233	226,6	95,2	1,1	465,6	206
<i>SPAC1B3.02c</i>	<i>SPAC1B3.02c</i>	mRNA	I	-	4928362	4928152	432,3	13,8	8,3	168,6	206
<i>SPAC29A4.15</i>	<i>SPAC29A4.15</i>	mRNA	I	-	5117470	5117409	2509,3	663,6	316,5	543,9	196
<i>SPBP8B7.05c</i>	<i>SPBP8B7.05c</i>	mRNA	II	-	3641071	3640938	259,7	19,4	609,8	1680,6	182
<i>mep33</i>	<i>SPBC28F2.02</i>	mRNA	II	+	1575649	1575836	127,4	15,7	0,6	376,6	179
<i>SPNCRNA.1511</i>	<i>SPNCRNA.1511</i>	ncRNA	II	-	2416444	2416358	2705,0	7440,7	224,4	2416,9	178
<i>rpl2102</i>	<i>SPAC959.08</i>	mRNA	I	+	3401081	3401127	1439,1	293,3	158,2	352,2	177
<i>edc3</i>	<i>SPBC18E5.11c</i>	mRNA	II	-	2093328	2092868	152,7	80,2	30,9	940,6	175
<i>rpl101</i>	<i>SPCC1183.08c</i>	mRNA	III	-	611852	611736	357,3	45,1	6,1	192,4	174
<i>SPBC1703.07</i>	<i>SPBC1703.07</i>	mRNA	II	+	2930353	2930402	590,5	85,2	341,8	620,4	174
<i>rpl35a</i>	<i>SPCP31B10.08c</i>	mRNA	III	-	540221	540141	1251,1	637,3	1,7	241,3	171
<i>ssc1</i>	<i>SPAC664.11</i>	mRNA	I	+	1728189	1728304	331,9	32,6	1,7	159,8	167
<i>def1</i>	<i>SPBC354.10</i>	mRNA	II	+	574209	574346	272,9	100,3	1,7	283,9	166
<i>SPBC660.16</i>	<i>SPBC660.16</i>	mRNA	II	+	233256	233345	159,3	206,2	5,0	941,8	166
<i>tif213</i>	<i>SPBC17G9.09</i>	mRNA	II	+	2189136	2189231	312,1	47,0	240,9	731,9	163
<i>lcf1</i>	<i>SPBC18H10.02</i>	mRNA	II	+	1772632	1772733	276,2	124,1	2,2	307,7	163
<i>eca39</i>	<i>SPBC428.02c</i>	mRNA	II	-	442470	442267	180,3	0,6	0,0	141,6	159
<i>rpl2501</i>	<i>SPBC106.18</i>	mRNA	II	+	414283	414503	1472,2	1094,1	17,6	340,9	156
<i>gpm1</i>	<i>SPAC26F1.06</i>	mRNA	I	-	5173912	5173864	4681,7	609,7	3983,1	1490,1	155
<i>SPAC1F8.07c</i>	<i>SPAC1F8.07c</i>	mRNA	I	-	101721	101524	367,8	424,2	0,6	439,3	154
<i>SPNCRNA.688</i>	<i>SPNCRNA.688</i>	ncRNA	I	-	819754	818900	484,7	174,8	8,8	236,2	153

^a La liste complète des gènes ayant un allongement de leur 3' UTR suite à la déplétion de Seb1 peut être obtenue sur demande.

^b RPM: Reads Per Million value based on 3' READS.

^c SS: Significance score for alternative polyadenylation (APA) regulation.

Formula:

$-S \cdot \log_{10}(p\text{-value})$; S = 1 for lengthened genes. For SS, +/- 999 indicate p -value beyond the limits of program.

Supplementary Table S4. List of the top 100 k-mer and their associated Z-score derived from pyMOTIF analysis of Seb1-HTP CRAC experiments.

Experiment #1		Experiment #2	
# of read clusters used to derived motif: 5913		# of read clusters used to derived motif: 9791	
k-mer length: min:4/max:8			
k-mer (motif)	Z-score	k-mer (motif)	Z-score
GUAG	20.01	GUAG	24.74
UGUA	17.27	UGUA	22.53
AGUA	16.90	GUAA	21.25
GUAA	16.55	UGUAG	20.20
AGUAA	15.23	AGUA	19.33
UGUAG	14.80	AGUAA	18.70
GUAAA	14.15	GUAAA	18.65
GUAGU	13.81	GUAGU	18.31
AGUAG	12.87	UGUAA	18.27
AGUAAA	12.46	UUGUA	17.86
UAGU	12.33	AGUAG	16.94

GUAGA	12.14	GUAGA	16.56
UGUAA	12.13	AGUAAA	15.70
UUGUA	12.08	UAGUA	15.69
GUAAU	12.05	UAGU	15.46
UAGUA	11.82	GUAAU	15.02
AAGUA	11.64	AAGUA	14.15
CGUA	11.05	AUGUA	13.81
AUGUA	11.01	CGUA	13.40
GUAU	9.89	UUGUAA	13.20
UCGUA	9.68	UGUAAA	13.04
AAAGUA	9.59	UGUAGU	12.99
GUAAAA	9.59	AAGUAA	12.66
GUAGC	9.42	GUAU	12.62
GUGUA	9.36	GUGUA	12.61
AGUAGU	9.33	UUGUAG	12.52
AGUAAU	9.31	AGUAGU	12.25
AAGUAA	9.11	GUGU	12.23
UAGUAA	9.02	CUGUA	12.19
AUGUAG	9.02	UAGUAA	11.80
UGUAU	8.90	GUAGC	11.73
UGUAGA	8.89	GUAAAA	11.65
CUGUA	8.71	GUAAAU	11.53
AGUAAAA	8.71	UGUAU	11.44
CGUAG	8.65	GUAGUU	11.44
UGUAGU	8.52	UUUGUA	11.43
AAGUAG	8.46	GUAGUA	11.39
UGUAAU	8.41	GUAAUU	11.29
GUAGUU	8.40	UGUAGA	11.23
UUGUAG	8.38	AGUAAU	11.23
GUAAAU	8.30	AUGUAG	10.96
UAGUAG	8.23	GUAAAAG	10.94
AAGUAAA	8.01	UGUGU	10.93
UGUAAA	8.01	GUAAG	10.90
AAAGU	7.98	AGUAGA	10.90
GUAAAAG	7.95	AUGUAA	10.82
GUGU	7.91	UUGU	10.72
AAGU	7.89	UCGUA	10.67
GAGUA	7.89	UAGUAG	10.61
UUUGUA	7.84	AAGUAG	10.57
GUAGG	7.84	AAAGUA	10.53
UUGU	7.82	AUUGUA	10.29
GUAGUA	7.74	UGUAAU	10.23
AGUAGA	7.67	GUAGAA	10.10
GUAAUU	7.67	GUGUAG	10.05
UUGUAA	7.66	GUAGAU	10.01

GUAUU	7.49	UGUGUA	9.87
AUGUAA	7.46	AAGUAAA	9.82
GUAGCA	7.32	GUAUU	9.80
GAGUAA	7.32	GUAGG	9.54
AGUAAAU	7.27	UGUAGC	9.53
UUGUAU	7.24	AAAGUAA	9.35
AGUG	7.24	CUUGUA	9.33
AUGU	7.23	AAUGUA	9.33
GUAGAU	7.21	CGUAG	9.31
CGUAA	7.18	GAGUA	9.30
CUGUAG	7.14	UGUG	9.27
GUAGAA	7.13	UAUGUA	9.23
GUAAAAU	6.99	UAGUAAA	9.14
CAGUA	6.98	AGUAAAA	9.05
AAUGUA	6.92	UUUGU	9.03
UUUUGUA	6.88	AGUAAAG	9.01
GUAAG	6.84	CUGUAG	8.95
AAAGUAA	6.84	GUGUG	8.91
GUAGUG	6.75	GUAAUG	8.84
GAGU	6.71	CGUAA	8.81
CAGUAA	6.70	GUUGUA	8.81
AGUGU	6.66	UUGUAU	8.80
UGUAGC	6.65	AGUAAAU	8.79
UAGUAAA	6.52	GUAGUG	8.76
UCGUAA	6.47	GUAAAAA	8.65
AUUGUA	6.47	UUGUAAA	8.62
GUAAAAA	6.47	GAGUAG	8.61
ACUGUA	6.43	UUUUGUA	8.58
AAGUAAAA	6.39	UGUAAG	8.45
UUAGUA	6.38	AGUG	8.44
GUAUG	6.31	UUUGUAA	8.43
GUAAAGU	6.26	GUAUG	8.39
UUGUAAU	6.24	UAGUG	8.38
AAAGUAG	6.22	AUAGUA	8.37
AGAGUA	6.22	GUAGUAA	8.30
UAUAGU	6.20	AUGUAAA	8.30
UAGUU	6.19	UGUAGG	8.29
GUGUG	6.17	UUAGUA	8.28
GUUG	6.15	CUGUAA	8.23
AUAGU	6.15	CUGU	8.22
GUAAAUU	6.13	UUGUAGU	8.21
AAUGU	6.12	GUAAAGU	8.13
AGUAU	6.10	GUAGCA	8.11
GUGUAA	6.10	GUAAAUU	8.00

Note : La liste complète des k-mer/Z-score peut être obtenue sur demande.

Supplementary Table S5. List of *S. pombe* strains used in this study.

Strain	Genotype	Reference
FBY106	h+ ade6-M210 leu1-32 ura4-D18 his3-D1	Perreault et al., 2007
FBY107	h- ade6-M216 leu1-32 ura4-D18 his3-D1 pab2::kanMX6	Perreault et al., 2007
FBY915	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6	This study
FBY954	h- ade6-M210 leu1-32 ura4-D18 his3-D1 Seb1-13XMyC::kanMX6	This study
FBY1250	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 snR99::ura4 pFB600::ade6	This study
FBY1256	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 snR99::ura4 pFB366::ade6	This study
FBY1298	h+ ade6-M21? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 pab2::natMX6	This study
FBY1464	h- Rpb3-3XHA::kanMX6	Coudreuse et al., 2010
FBY1496	h? ade6? leu1-32 ura4-D18 his3-D1 nab3::kanMX6	This study
FBY1507	h? ade6? leu1-32 ura4-D18 his3-D1 sen1::kanMX6	This study
FBY1516	h? ade6? leu1-32 ura4-D18 his3-D1 dbl8::kanMX6	This study
FBY1560	h+ ade6-M210 leu1-32 ura4-D18 his3-D1 pFB366::ade6	This study
FBY1561	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 pFB366::ade6	This study
FBY1622	h? ade6? leu1-32 ura4-D18 his3-D1 dbl8::kanMX6 sen1::natMX6	This study
FBY1712	h+ ade6-M210 leu1-32 ura4-D18 his7-366 pFB45::ura4	This study
FBY1827	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 pFB941::ade6	This study
FBY1829	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 pFB943::ade6	This study
FBY1831	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 pFB944::ade6	This study
FBY1833	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 pFB946::ade6	This study
FBY1858	h+ ade6-M210 leu1-32 ura4-D18 his3-D1 Seb1-HTP::kanMX6	This study
FBY1920	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 pFB1037::ade6	This study
FBY1950	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 Rna14-4XTAP::natMX6	This study
FBY1953	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 Rna14-4XTAP::natMX6	This study
FBY1956	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 Ysh1-4XTAP::natMX6	This study
FBY1959	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 Ysh1- 4XTAP::natMX6	This study
FBY1962	h- Rpb3-3XHA::kanMX6 P41nmt1-seb1::natMX6	This study
FBY1979	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 Clp1(SPAC22H10.05c)- 4XTAP::natMX6	This study
FBY1980	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 Clp1(SPAC22H10.05c)-4XTAP::natMX6	This study
FBY1981	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 Cft2-4XTAP::natMX6	This study
FBY1982	h+ ade6? leu1-32 ura4-D18 his3-D1 P41nmt1-seb1::kanMX6 Cft2- 4XTAP::natMX6	This study
FBY2007	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 snR99::ura4 pFB1057::ade6	This study
FBY2009	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 snR99::ura4 pFB1059::ade6	This study

FBY2011	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 snR99::ura4 pFB1061::ade6	This study
FBY2013	h- ade6-M210 leu1-32 ura4-D18 his3-D1 Seb1-13XMyC::kanMX6 Cft2-4XTAP::natMX6	This study
FBY2015	h- ade6-M210 leu1-32 ura4-D18 his3-D1 Seb1-13XMyC::kanMX6 Ysh1-4XTAP::natMX6	This study
FBY2063	h+ ade6-M210 leu1-32 ura4-D18 his7-366 pFB45::ura4 P41nmt1- seb1::natMX6	This study

Supplementary Table S6. List of *S. cerevisiae* strains used in this study.

Strain	Genotype	Reference
FBY1801	MATa ura3-1 ade2-1 his3-11,5 trp1-1 leu2-3,112 can1-100	Thomas et Rothstein. 1989
FBY1802	MATa ura3-1 ade2-1 his3-11,5 trp1-1 leu2-3,112 can1-100 <HIS>PGAL1::3XHA-NRD1	Thiebaut et al.. 2006
FBY1899	MATa ura3-1 ade2-1 his3-11,5 trp1-1 leu2-3,112 can1-100 [pFB950::LEU2]	This study
FBY1900	MATa ura3-1 ade2-1 his3-11,5 trp1-1 leu2-3,112 can1-100 <HIS>P _{GAL1} ::3XHA-NRD1 [pFB950::LEU2]	This study
FBY1901	MATa ura3-1 ade2-1 his3-11,5 trp1-1 leu2-3,112 can1-100 <HIS>P _{GAL1} ::3XHA-NRD1 [pFB974::LEU2]	This study
FBY1903	MATa ura3-1 ade2-1 his3-11,5 trp1-1 leu2-3,112 can1-100 <HIS>P _{GAL1} ::3XHA-NRD1 [pFB1007::LEU2]	This study

Supplementary Table S7. List of plasmids used in this study.

Strain	Genotype	Reference
pFB45	Empty <i>ura4</i> chromosomal integration vector (pJK210)	Keeney & Boeke. 1994
pFB366	Empty <i>ade6</i> chromosomal integration vector	Beaudoin et al. 2006
pFB600	snR99 noncoding sequence flanked by snR99 promoter and terminator sequences (WT)	This study
pFB941	nrd1-3XFLAG in pFB366 flanked by seb1 promoter and terminator sequences	This study
pFB943	seb1-3XFLAG in pFB366 flanked by seb1 promoter and terminator sequences	This study
pFB944	pFB943 with C-terminal-interacting domain mutations (CID _{mut}) (Y64A/D67A/R71A)	This study
pFB946	pFB943 with RRM mutations (RRM _{mut}) (R442A/H443A/F445A/K447A)	This study
pFB950	<i>S. cerevisiae</i> centromeric plasmid (pRS315, LEU2, CEN)	Sikorski & Hieter, 1989
pFB971	Plasmid to amplify HTP-tagging cassette	From S. Granneman.
pFB974	nrd1-3XFLAG in pFB950 flanked by nrd1 promoter and terminator sequences	This study
pFB1007	seb1-3XFLAG (without intron) in pFB950 flanked by nrd1 promoter and terminator sequences	This study
pFB1037	pFB943 with Ser ⁵ _{mut} mutations (S22A/K25A)	This study
pFB1057	pFB600 with deletion in terminator sequences (Mutant #1)	This study
pFB1059	pFB600 with deletion in terminator sequences (Mutant #2)	This study
pFB1061	pFB600 with deletion in terminator sequences (Mutant #1/2)	This study

ARTICLE 3**THE RNA EXOSOME PROMOTES TRANSCRIPTION TERMINATION OF
BACKTRACKED RNA POLYMERASE II**

Jean-François Lemay^{1, 4}, Marc Larochelle^{1, 4}, Samuel Marguerat², Sophie Atkinson³,
Jürg Bähler³, and François Bachand¹

Publié dans: Nature Structural & Molecular Biology (2014) 21: 919-926.

¹ RNA Group, Department of Biochemistry, Université de Sherbrooke, Sherbrooke, Québec, Canada.

² Medical Research Council Clinical Sciences Centre, Imperial College London, London, U.K.

³ Research Department of Genetics, Evolution and Environment, University College London, London, U.K.

⁴ These authors contributed equally to this work.

CONTRIBUTION

Cet article est le résultat d'une coopération entre les laboratoires des professeurs Marguerat, Bähler et Bachand.

Pour ma part, j'ai créé la quasi-totalité du matériel biologique utilisé lors de cette étude. J'ai notamment généré le système rapporteur snR99T-ura4, un système utilisé pour établir la présence de défauts de terminaison transcriptionnelle lorsque le niveau protéique de l'exosome est sous-optimal. J'ai également créé les systèmes permettant l'expression de versions mutantes de Dis3 et de Rrp41 afin d'établir la contribution enzymatique et l'importance du canal central de l'exosome dans la terminaison de la transcription. De plus, j'ai conçu le système rapporteur *ura4* afin de tester l'influence d'éléments en *cis* associés à la pause transcriptionnelle sur la capacité de l'exosome à ARN à promouvoir la terminaison de la transcription. Finalement, j'ai généré le matériel nécessaire afin d'établir le rôle du *backtracking* de l'ARNPII dans le processus de terminaison de la transcription exosome à ARN-dépendant. Globalement, j'ai donc contribué amplement et significativement, d'un point de vue technique et intellectuel, à l'aboutissement de cette étude. Sommaire des figures/expériences effectuées par J-F.L. : Fig. 1A, D; Fig. 5A-F; Fig. 6D, F; Fig. S1B-D; Fig. S2C; Fig. S3; Fig. S4C-H; Fig. S5; Fig. S6; Fig. S7.

S.M. et S.A., sous la supervision de J.B., ont préparé les bibliothèques associées aux expériences de RNA-seq. S.M. a de plus effectué l'ensemble des analyses bio-informatiques de ce papier. M.L. est responsable des expériences d'inactivation de l'exosome suite à stress thermique et de plusieurs expériences de ChIP. Celles-ci incluent les ChIP dédiées au séquençage à haut débit (ChIP-seq) ainsi que les ChIP contre les sous-unités de l'exosome (Dis3-TAP, Rrp41-TAP, Csl4-TAP, Rrp4-TAP) et le facteur de maturation 3' de l'ARN Pcf11. M.L. a, de plus, été responsable de l'exécution des expériences de TRO.

RÉSUMÉ

L'exosome à ARN est un complexe enzymatique dont le rôle fondamental consiste à assurer une expression génique optimale notamment en limitant post-transcriptionnellement la production de transcrits aberrants. La surveillance d'ARN par l'exosome ne semble toutefois pas uniquement restreinte au niveau post-transcriptionnel, de nombreuses études ayant rapportées une certaine promiscuité entre l'exosome et le processus transcriptionnel. Cependant, les détails mécanistiques sous-jacents à cette relation sont méconnus. Dans cette étude, nous présentons des données indiquant que l'exosome, de l'organisme modèle *S. pombe*, intervient au niveau de la transcription en favorisant la terminaison d'évènements transcriptionnels associés à des ARNPIIs ayant pausées et reculées le long de la matrice d'ADN. Nos résultats supportent un mécanisme, qui va à l'encontre du concept présumé voulant que l'exosome à ARN agisse exclusivement qu'au niveau d'ARNs libérés de l'ARNPII, où l'exosome à ARN cible pour dégradation une extrémité 3' d'ARN nouvellement accessible suite au *backtracking* des ARNPIIs afin de promouvoir la terminaison de la transcription. Ces découvertes révèlent ainsi un mécanisme co-transcriptionnel de surveillance d'ARN où la terminaison de la transcription par l'exosome à ARN prévient notamment l'accumulation d'ARN aberrants et l'interférence transcriptionnelle entre gènes adjacents.

ABSTRACT

The exosome is an RNA decay complex that constantly scrutinizes transcription and contributes to post-transcriptional turnover of faulty mRNAs. Yet, how nuclear RNA surveillance by the exosome is coordinated with transcription is still unknown. Here we show that the RNA exosome of *Schizosaccharomyces pombe* can target the transcription machinery by terminating transcription events associated with paused and backtracked RNA polymerase II (RNAPII), contrary to the notion that the exosome acts exclusively on RNAs that have been released by RNAPII. Our data support a mechanism by which RNAPII backtracking provides a free RNA 3' end for the core exosome, resulting in transcription termination with the concomitant degradation of the associated transcript. These findings uncover a mechanism of co-transcriptional RNA surveillance whereby termination of transcription by the exosome prevents aberrant read-through RNAs and transcriptional interference at neighboring genes.

INTRODUCTION

Gene transcription by RNA polymerase II (RNAPII) produces a variety of functional transcripts that include messenger RNAs (mRNA), small nucleolar RNAs (snoRNA), and small nuclear RNAs (snRNA). For a majority of RNAPII-transcribed gene, RNA synthesis involves extensive processing events that occur simultaneously with their synthesis (Bentley, 2014). In the case of mRNAs, a nascent pre-mRNA will be processed co-transcriptionally by the addition of a 7-methylguanosine cap at the 5' end, intron removal, 3' end endonucleolytic cleavage, and polyadenylation. One noteworthy example of coordination between RNA processing and transcription occurs during 3' end maturation of nascent transcripts, which is coupled to transcription termination (Bentley, 2014, Mischo et Proudfoot, 2013). Current data suggest a model whereby the co-transcriptional transfer of mRNA 3' end processing factors from RNAPII onto the nascent pre-mRNA promotes cleavage at the poly(A) site, providing a free 5' end entry point for the 5'-3' exonuclease Rat1 (Xrn2), which is thought to chase RNAPII and promote its dissociation from the DNA template by a mechanism that remains unclear (Dengl et Cramer, 2009, Jimeno-Gonzalez *et al.*, 2014, Kim *et al.*, 2004, Luo *et al.*, 2006, Pearson et Moore, 2013, West *et al.*, 2004). Although this “torpedo” model of transcription termination depends on prior cleavage of the nascent RNA by 3' end processing factors, there is also evidence to support that the transcription process reciprocally influences 3' end processing. Indeed, RNAPII binding studies show that RNAPII tend to accumulate at the 3' end of genes (Anamika *et al.*, 2012, Coudreuse *et al.*, 2010, Glover-Cutter *et al.*, 2008, Gromak *et al.*, 2006), suggesting that transcription pauses at or near the poly(A) site to facilitate cleavage site recognition by the 3' end processing factors.

The aforementioned connection between transcription and RNA processing is thought to promote efficient RNA maturation; yet, it is also subject to inaccurate and unproductive processing events that lead to the expression of defective transcripts. As the number of mRNA molecules per cell are often low (Marguerat *et al.*, 2012), it is clear that errors anywhere in the gene expression process could have profound consequences on cellular functions. To limit the production of aberrant transcripts, the cell uses a range of surveillance mechanisms, among which is the RNA exosome. The exosome is a conserved RNA

degradation machine that plays a fundamental role in monitoring the quality of gene expression in the nucleus. In eukaryotes, the exosome complex is minimally composed of ten evolutionarily conserved subunits, including the catalytically active ribonuclease, Dis3 (Rrp44), which exhibits both 3'-5' exonuclease and endonuclease activities (Chlebowski *et al.*, 2013, Schneider et Tollervey, 2013). In the yeast nucleus, the core exosome is associated with an additional 3'-5' exonuclease, Rrp6 (Chlebowski *et al.*, 2013). Although early studies suggested that Dis3 and Rrp6 were functionally redundant, the emergence of genetic and biochemical evidence indicate that they are likely to have different cellular activities (Januszyk et Lima, 2014).

As for RNA maturation, circumstantial evidence suggest that RNA surveillance by the nuclear exosome is also connected to the transcription process (Schmid et Jensen, 2013). The nuclear exosome in *Drosophila* has been shown to associate directly with the transcription elongation complex (Andrulis *et al.*, 2002). Studies in *Saccharomyces*, *Chironomus*, *Drosophila*, and humans indicate that the RNA exosome is present at actively transcribed genes (Basu *et al.*, 2011, Hesse *et al.*, 2009, Hieronymus *et al.*, 2004, Lim *et al.*, 2013). Recent results also suggest that the exosome can influence the transcription process (Castelnuovo *et al.*, 2013, Shah *et al.*, 2014, Wagschal *et al.*, 2012). Accordingly, the current view is that the nuclear exosome constantly scrutinize transcription events and rapidly degrades aberrant RNAs post-transcriptionally. As yet, however, the mechanistic understanding of how the nuclear exosome monitors RNA processing during transcription remains elusive.

To examine this question, we depleted core exosome subunits in the fission yeast *S. pombe* and analyzed the polyadenylated transcriptome and the occupancy profile of RNAPII on a genome-wide scale. We show that read-through transcription from many RNAPII-transcribed genes into adjacent genes is prevented by the 3'-5' exonucleolytic activity of the core exosome. Furthermore, we show that transcription termination defects in exosome-deficient cells depend on TFIIIS, a transcription factor that resumes transcription elongation of paused and backtracked RNAPII. Our results support a mechanism of transcription termination whereby RNAPII backtracking at pause sites exposes an unprotected RNA 3'

end that can be targeted by the 3'-5' exonucleolytic activity of the RNA exosome to promote termination. These findings provide new insights into co-transcriptional RNA surveillance by uncovering a function for the core exosome in a transcription termination pathway that is coupled to RNA decay.

RESULTS

Accumulation of 3'-extended RNAs in exosome-deficient cells

We constructed strains with endogenous *dis3* and *rrp41* genes, which encode core exosome subunits, under the control of the thiamine-repressible *nmt1* promoter (P_{nmt1}). Growing P_{nmt1} -*dis3* cells in thiamine-supplemented medium depleted Dis3 by over 85%, resulting in a severe processing defect of a known rRNA substrate (Supplementary Fig. 1a-b). Notably, depletion of core exosome subunits resulted in the production of 3'-extended mRNAs (Fig. 1a, lanes 4 and 6) and snoRNAs (Supplementary Fig. 1c-d). In contrast, levels of regular length mRNA were not reduced (Fig. 1a), arguing against cleavage defects by the 3' end processing machinery. Analysis of Pcf11 recruitment further supports normal 3' end processing in exosome-deficient cells (Supplementary Fig. 1e). Moreover, the 3'-extended mRNAs detected in exosome-depleted cells were cleaved and polyadenylated downstream of *cis* sequences appropriate for 3' end processing (Supplementary Fig. 2a-b). 3'-extensions could also be detected just 15 min after exosome inactivation (Fig. 1b) using a previously described temperature-sensitive strain for *dis3* (Yamanaka *et al.*, 2010), suggesting that the 3'-extended transcripts that accumulate in *dis3* and *rrp41* conditional mutants are a direct result of exosome depletion. The RNA degradation function of the nuclear exosome can be promoted by the activity of an evolutionarily conserved polyadenylation complex, called TRAMP, which in *S. pombe* consist of the RNA helicase Mtr4, the poly(A) polymerase Cid14, and the zinc knuckle RNA-binding protein Air1 (Keller *et al.*, 2010). Interestingly, read-through transcripts were not detected in cells deficient for components of the TRAMP

complex (Fig. 1c-d) and were also much less evident in a *rrp6Δ* mutant (Supplementary Fig. 3), suggesting a more prominent role for the core exosome relative to the exosome-associated exonuclease, Rrp6.

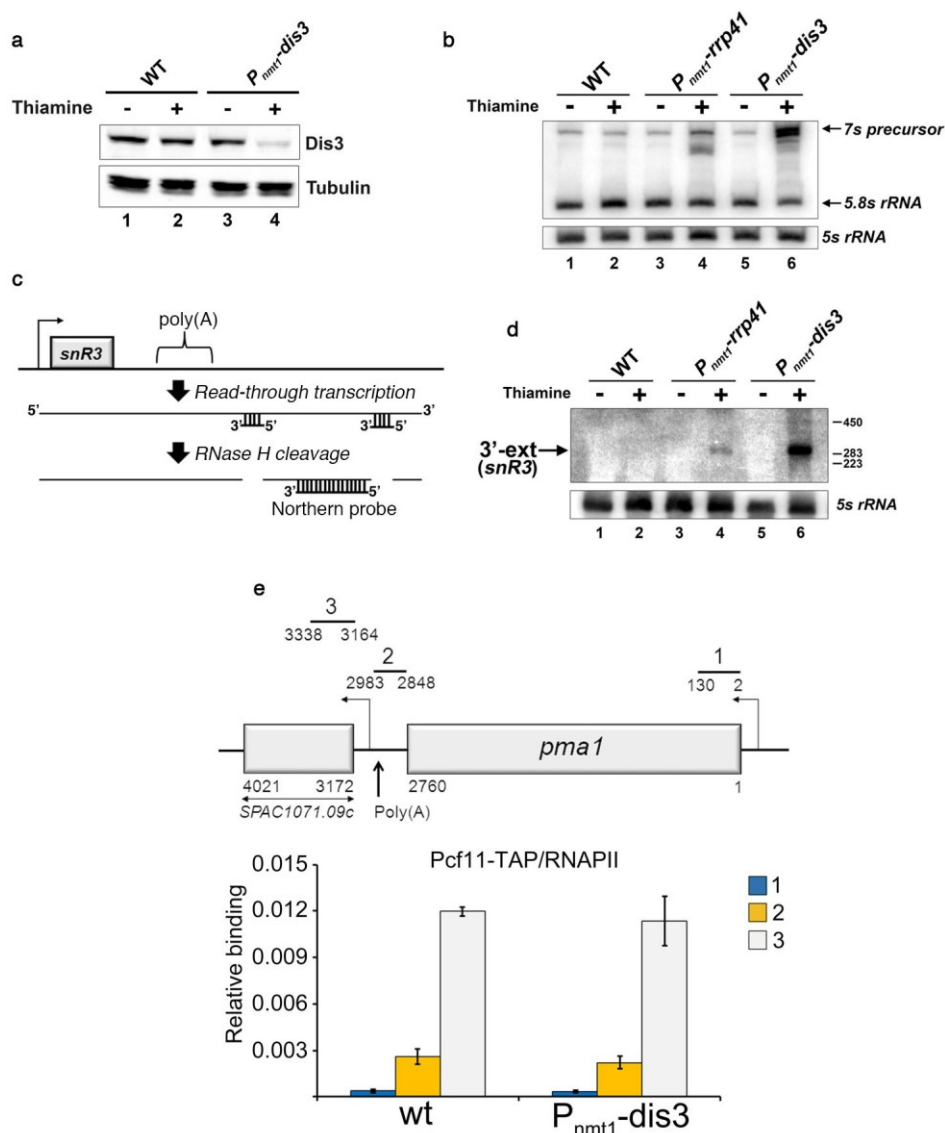


Figure S1. Exosome depletion results in the production of read-through RNAs.

(a) Western blot analysis of extracts prepared from wild-type (lanes 1-2) and *P_{nmt1}-dis3* (lanes 3-4) cells that were grown in the absence (-) or presence (+) of thiamine. (b) Northern blot analysis of RNA prepared from wild-type (lanes 1-2), *P_{nmt1}-rrp41* (lanes 3-4), and *P_{nmt1}-dis3* (lanes 5-6) cells that were grown in the absence (-) or presence (+) of thiamine. Note the thiamine-dependent increase in 5.8S rRNA precursor (7S rRNA) and concomitant decrease in mature 5.8S rRNA in Dis3- and Rrp41-depleted cells (lanes 4 and 6, respectively), but not in thiamine-treated wild-type cells (lane 2). (c) Schematic of the RNase H cleavage assay used to detect 3'-extension of the snR3 snoRNA. (d) Equal amounts of total RNA prepared from the indicated strains grown in the absence (-) or in the presence (+) of thiamine were

treated with RNase H in the presence of DNA oligonucleotides complementary to sequences located beyond the polyadenylation site (poly(A)) of snR3. The positions of the DNA oligonucleotides used are shown in the schematic representation (c). Read-through snR3 transcripts (3'-ext) were detected using a strand-specific riboprobe complementary to the RNase H-cleaved product. Size markers (nt) are indicated on the right. The 5S rRNA was used as a loading control. (e) Binding of Pcf11-TAP relative to RNAPII occupancy was calculated for three independent regions of the *pml* gene (shown on the schematic above the graph) in wild-type (WT) and P_{nmt1} -dis3 cells grown in the presence of thiamine. Percent inputs for Pcf11-TAP were corrected for Pol II occupancy along *pml* using the CTD-specific antibody (8WG16). Bars indicate the range of measured values from two independent experiments.

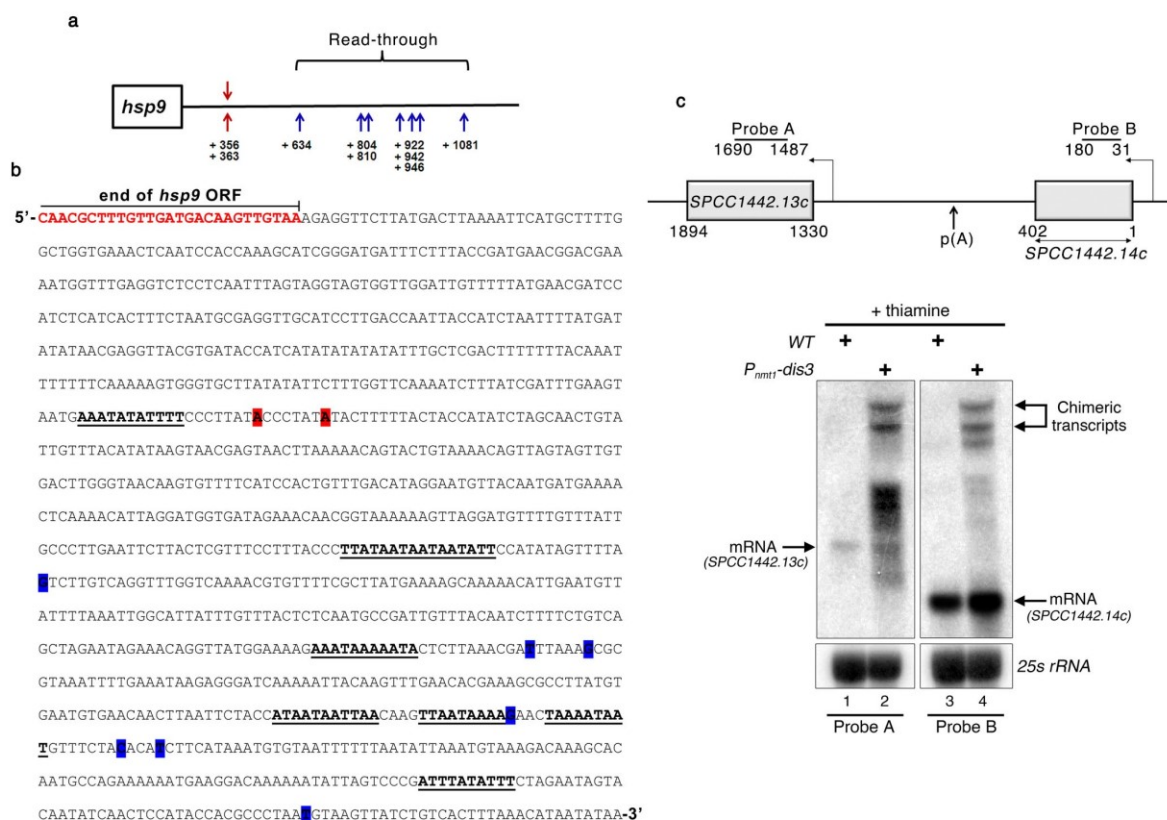


Figure S2. Read-through RNAs detected in exosome-depleted cells are cleaved and polyadenylated downstream of motifs associated with 3' end processing and can result in the production of chimeric transcripts.

(a) Schematic summarizing oligo d(T)-primed 3' RACE analyses of polyadenylation site positions for *hsp9* in wild-type and Dis3-depleted cells. Red arrows indicate the position of polyadenylation sites for the regular length *hsp9* mRNA in both wild-type and Dis3-depleted cells, whereas blue arrows indicate the polyadenylation sites for the read-through *hsp9* mRNAs detected in Dis3-depleted cells. (b) Shown are the nucleotide positions of poly(A) sites identified by 3' RACE for regular length (red) and read-through (blue) mRNAs in the *hsp9* gene. The underlined DNA sequences represent motifs known to be over-represented

in regions -10 to -30 relative to cleavage site^{61,62}, including the AATAAA hexamer. (c) Northern analysis of total RNA prepared from wild-type (lanes 1 and 3) and *P_{nmt1}-dis3* (lanes 2 and 4) cells that were supplemented with thiamine. The blot was hybridized with an antisense RNA probe complementary to *SPCC1442.13c* (lanes 1-2; Probe A) and *SPCC1442.14c* (lanes 3-4; Probe B). The position of the *SPCC1442.14c* poly(A) site was identified by RNA-seq⁶². Note that the *SPCC1442.14c-SPCC1442.13c* chimeric transcripts are recognized using both RNA probes. The 25S rRNA was used as a loading control.

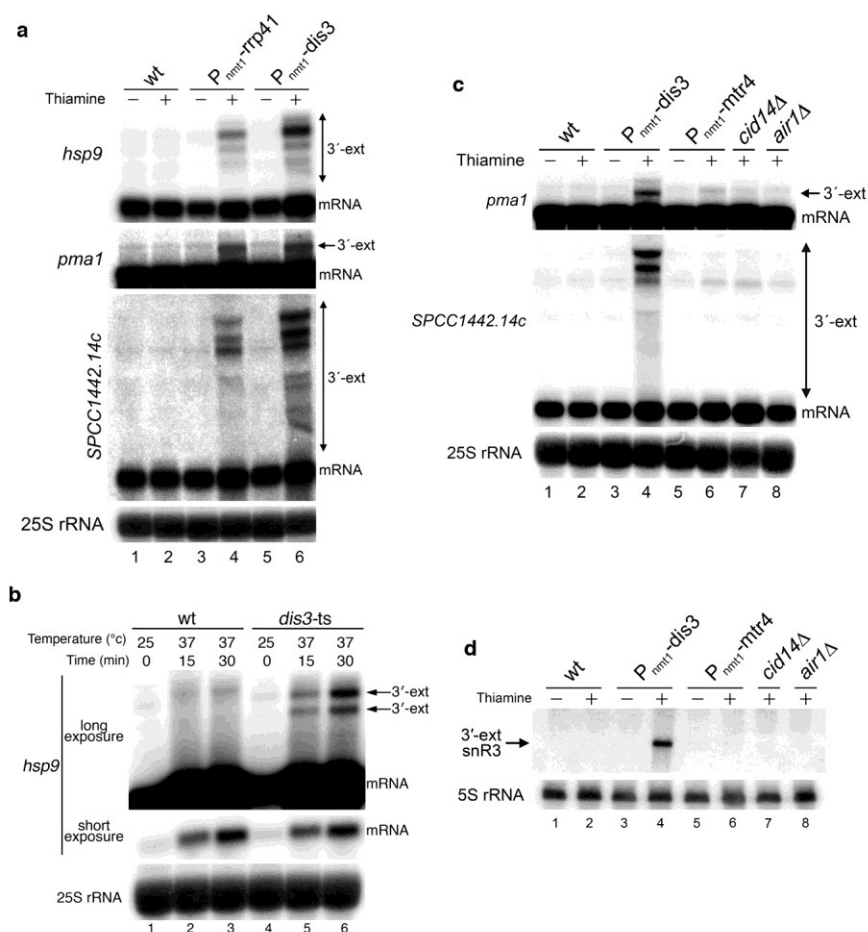


Figure 1. Cells deficient in core exosome subunits, but not the TRAMP complex, accumulate read-through RNAs.

(a) Northern analysis of RNA from the indicated strains grown in the absence (-) or presence (+) of thiamine. wt, wild-type. 3'-ext, 3'-extended transcripts. (b) Northern analysis of *hsp9* mRNA prepared from wild-type (lanes 1-3) and *dis3-ts* (lanes 4-6) cells that were grown at the permissive temperature (lanes 1 and 4) or that were shifted to the non-permissive temperature for 15 min (lanes 2 and 5) or 30 min (lanes 3 and 6). (c) Northern analysis of RNA from the indicated strains grown in the absence (-) or presence (+) of thiamine. (d) Equal amounts of total RNA extracted from the indicated strains grown in the absence (-) or in the presence (+) of thiamine were treated with RNase H in the presence of DNA

oligonucleotides complementary to snR3 transcripts present beyond the poly(A) site. The position of the DNA oligonucleotides used are shown in Supplementary Fig. 1c. 3'-extended snR3 transcripts (3'-ext) were detected using a RNA probe complementary to the RNase H-cleaved product.

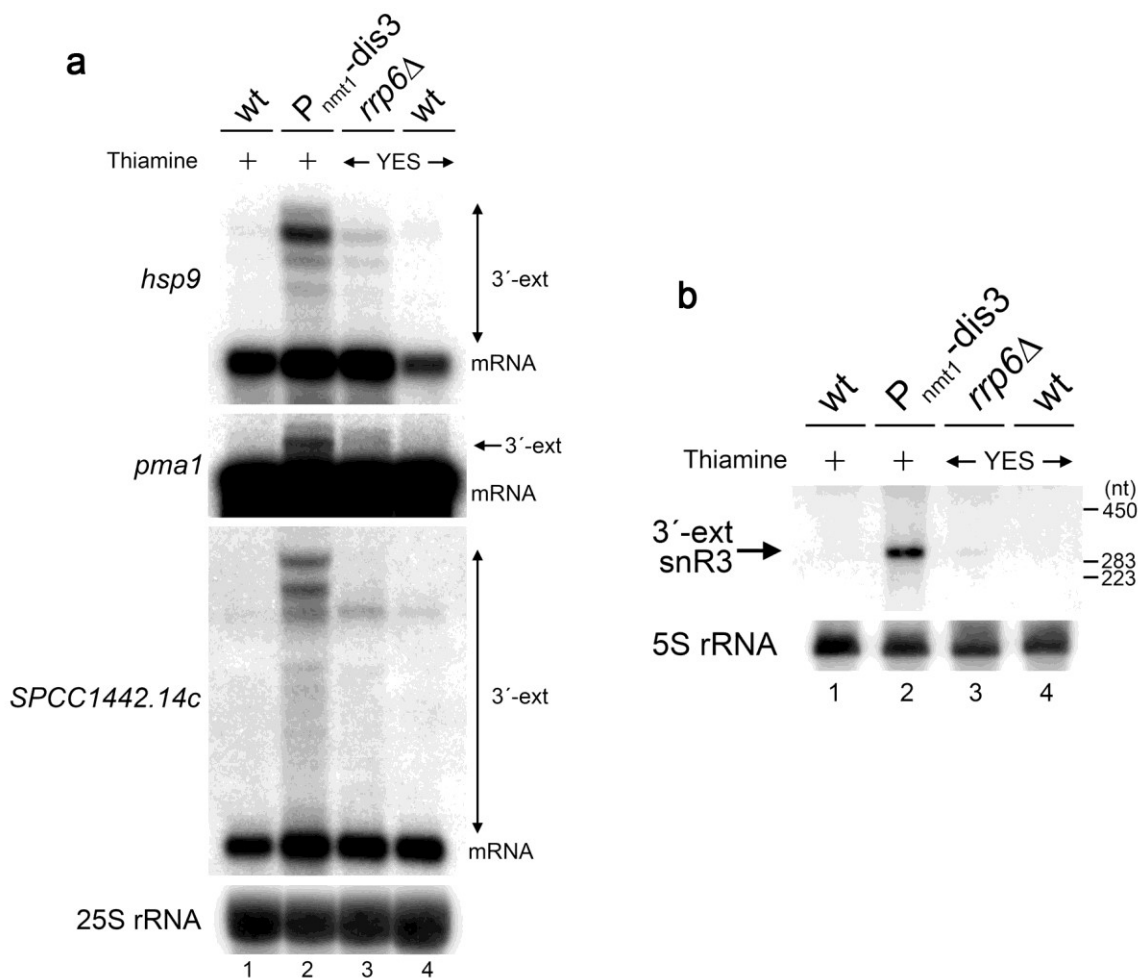


Figure S3. The Rrp6 exoribonuclease does not extensively contribute to the suppression of read-through transcript.

(a) Northern blot analysis of total RNA prepared from wild-type (lanes 1 and 4), P_{nmt1} -*dis3* (lane 2), and *rrp6* Δ (lane 3) cells that were grown in minimal media supplemented with thiamine (lanes 1-2) or rich media (YES; lanes 3-4). The blots were analyzed using strand-specific antisense RNA probes. The 25S rRNA was used as a loading control. (b) Northern analysis of RNase H-cleaved *snR3* 3'-extensions as described in Supplementary Fig. 1c using total RNA prepared from cells described in (a).

To measure the extent of read-through RNA production in exosome-deficient cells, we analyzed the polyadenylated transcriptome of wild-type and mutant cells by strand-

specific RNA sequencing. Consistent with our northern analyses, exosome depletion led to the detection of 3'-extended transcripts expressed from coding and noncoding genes that were not observed in Mtr4-depleted cells (Fig. 2a). Out of 5203 genes encoding mRNAs, snoRNAs, and snRNAs, 1692 (33%) and 1562 (30%) genes showed a significant increase of at least two-fold in the accumulation of 3'-extended transcripts in Dis3- and Rrp41-depleted cells, respectively, with >70% of these genes affected in both mutants ($p < 2.2e-16$, one-sided Fisher's exact test). The frequency of read-through RNA production in exosome mutants correlated positively to gene expression levels (Fig. 2b). The prevalence of 3'-extended transcripts in exosome-deficient cells as identified by RNA-seq can be visualized by plotting cumulative transcript level on the same strand before and after polyadenylation sites (Fig. 2c): whereas this level drops substantially after polyadenylation sites in RNA from wild-type and Mtr4-depleted cells, it remains high in Dis3- and Rrp41-depleted cells. Consistent with this observation, the distribution of mapped cleavage sites indicated that distal poly(A) sites were more frequently used in exosome mutants relative to wild-type cells (Fig. 2d). Collectively, these results indicate that loss of exosome function results in the widespread production of 3'-extended transcripts.

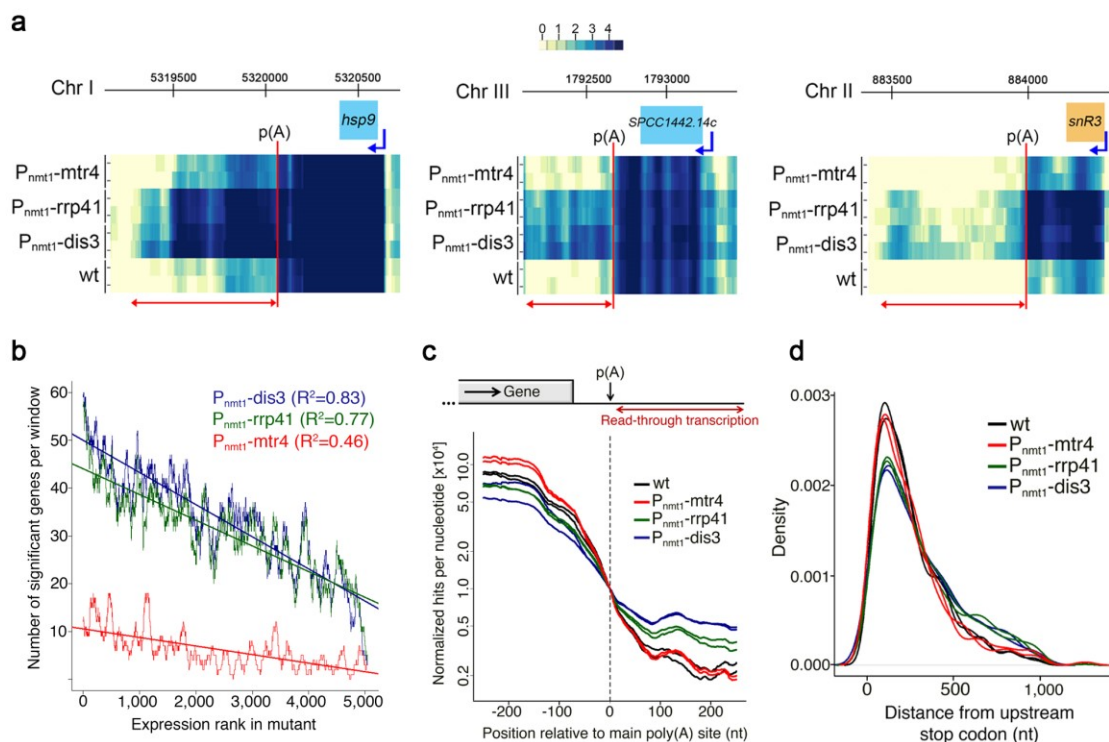


Figure 2. Widespread production of 3'-extended transcripts in exosome-depleted cells. (a) Heatmaps of duplicate data sets showing normalized RNA-seq scores along three genes for the indicated strains grown in the presence of thiamine. Note the increased density of mapped reads downstream of annotated poly(A) sites (red vertical line). For *snR3*, p(A) refers to the polyadenylation site of its 3'-extended precursor. Transcription start sites are indicated by arrows. Chr, Chromosome. (b) 5203 genes encoding mRNAs, snoRNAs, and snRNAs were ranked in decreasing order according to their expression levels (*x*-axis, RPKM levels measured by RNA-seq). The number of genes showing significant read-through in a window of 100 genes downstream of each position is represented on the *y*-axis. Linear regression on data from each mutant is shown as a plain line. Coefficient of determination (R^2) is shown. (c) Cumulative RNA read-through in the indicated strains. Curves show the sum of normalized RNA-seq scores for duplicate data sets over a genomic region covering the main poly(A) site. (d) Distribution of RNA-seq-derived cleavage site (CS) for duplicate data sets from the indicated strains with respect to the upstream stop codon. Mapped CS in exosome mutants are on average significantly more distant from the stop codon as compared to wild-type cells ($p < 3.2e-5$ and $p < 2.2e-3$ for *dis3*; $p < 1.1e-4$ and $p < 4.9e-6$ for *rrp41*; Wilcoxon rank sum test). No significant change was found for Mtr4-depleted cells. Experiments were performed in biological duplicates (independent cell cultures).

RNAPII termination defects in exosome-depleted cells

At least two different mechanisms could account for the increased production of 3'-extended transcripts in exosome-deficient cells: post-transcriptional turnover of aberrant read-through RNAs (Vasiljeva et Buratowski, 2006, West *et al.*, 2006) and/or a role for the exosome in transcription termination. If all of the 3'-extended transcripts detected in exosome-deficient cells originate from the stabilization of constitutively transcribed read-through RNAs, no change in RNAPII occupation across genes is expected following loss of exosome function. Yet, if the core exosome is involved in transcription termination, increased levels of RNAPII density should be detected at the 3' end of genes. To measure RNAPII density along genes, chromatin immunoprecipitation (ChIP) assays were performed using an antibody (8WG16) that recognizes the carboxy-terminal domain (CTD) of RNAPII, and analyzed by high-throughput sequencing (ChIP-seq). Analysis of RNAPII density at a snoRNA (Fig. 3a) and a mRNA (Fig. 3b) gene that showed accumulation of 3'-extended transcripts in exosome mutants (Figs. 1-2) revealed greater RNAPII crosslinking at the 3' end in Dis3- and Rrp41-depleted cells relative to the wild-type. Accordingly, transcription through the *SPCC1442.14c* terminator region (Fig. 3b) resulted in the production of chimeric *SPCC1442.14c-SPCC1442.13c* transcripts in exosome-deficient cells (Supplementary Fig. 2c). Importantly, transcription termination defects are a general feature of exosome-depleted cells, as seen by plotting cumulative levels of RNAPII binding relative to annotated cleavage sites, which indicated a clear shift in RNAPII density downstream of the noticeable decline observed in wild-type cells (Fig. 3c). The genome-wide increase in read-through transcription was also detected by a CTD-independent ChIP approach that used an antibody to the HA-tag of a core RNAPII component (Rpb3-HA; Fig. 3d). Several classes of RNAPII-transcribed genes showed a 3'-extended profile of RNAPII crosslinking in exosome mutants, including mRNA-, snoRNA-, and snRNA-encoding genes (Supplementary Fig. 4a) as well as both tandem and convergent gene pairs (Supplementary Fig. 4b). Exosome-dependent transcription termination defects were also detected using a reporter gene assay in which the transcription termination element of the *snR99* gene was inserted upstream of the *ura4* gene (Supplementary Fig. 4c-h). We also measured the distribution of actively elongating RNAPII by transcription run-on (TRO) assays on the *pma1* (Fig. 3e) and *snR3* (Fig. 3h) loci. In agreement with the ChIP-seq data, TRO assays revealed increased production of nascent RNA at the 3' end in Dis3- and Rrp41-depleted cells (Fig. 3f and 3i). Therefore, a significant

proportion of RNAPII that terminate downstream of *pma1* and *snR3* coding regions in wild-type cells fail to terminate in exosome mutants (Fig. 3g and 3j), indicating that the read-through transcripts detected in exosome mutants are caused by a termination defect and not simply by RNA stabilization. Collectively, the widespread detection of 3'-extended transcripts in exosome-deficient cells as determined by RNA-sequencing together with the global increase in RNAPII density at the 3' end of genes strongly support the idea that the exosome functions in RNAPII termination.

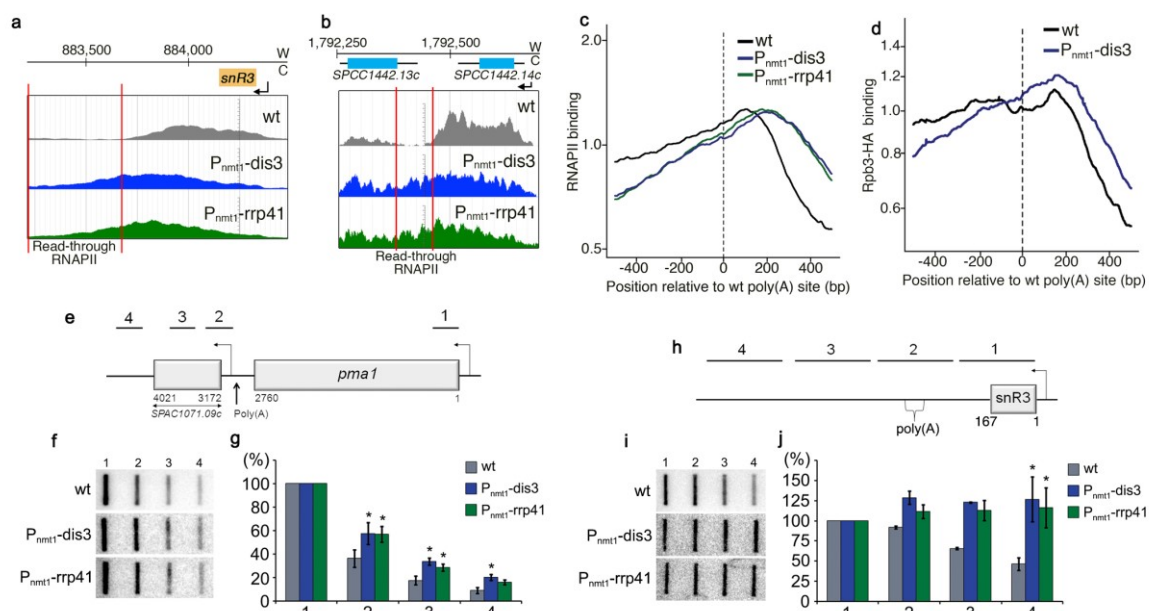


Figure 3. The core exosome promotes termination of RNAPII transcription.

(a-b) RNAPII profiles (ChIP-seq) across the *snR3* (a) and *SPCC1442.14c* (b) genes in the indicated strains. W, Watson strand; C, Crick strand. (c-d) Cumulative Rpb1 (c) and Rpb3-HA (d) profiles relative to poly(A) cleavage sites in the indicated strains. Curves show the sum of normalized ChIP-seq sequencing scores over a genomic region covering the main poly(A) site. (e-j) Schematic showing the position of probes used for transcription run-on (TRO) assays on *pma1* (e) and *snR3* (h) genes. Representative TRO blots are shown for *pma1* (f) and *snR3* (i). Quantification of TRO analyses for *pma1* (g) and *snR3* (j) genes. Transcription levels were corrected for background signal and plotted relative to probe 1. (g) Error bars, s.d. (n = 3 biological replicates from independent cell cultures). * $P < 0.05$ (two-tailed Student's *t* test). (j) Bars indicate either the range of measured values from two independent experiments (probes 2-3) or the s.d. (probes 1 and 4; n = 3 biological replicates from independent cell cultures). * $P < 0.05$ (two-tailed Student's *t* test).

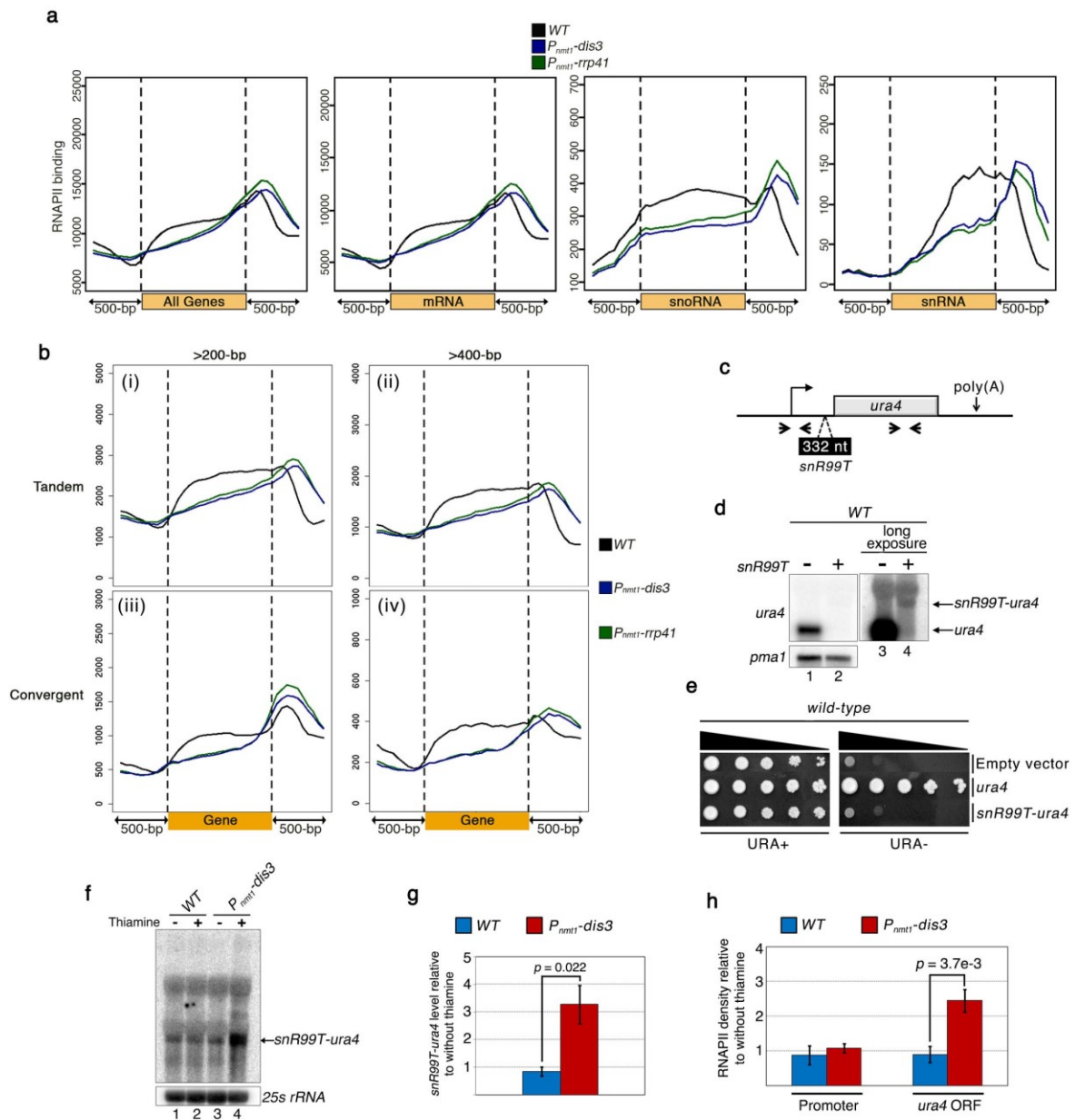


Figure S4. Read-through transcription by RNAPII in exosome-deficient cells.

(a) Average distribution of RNAPII density in the indicated strains, as determined by analysis of ChIP-seq results for all ($n=7005$), mRNA ($n=5123$), snoRNA ($n=53$), and snRNA ($n=7$) genes. Orange rectangles represent open reading frames (mRNA genes) and coding sequences (snoRNA and snRNA genes) and black arrows 500-bp of upstream and downstream flanking DNA sequences. (b) Average distribution of RNAPII density in the indicated strains, as determined by analysis of ChIP-seq results for (i) tandem genes that are more than 200-bp apart ($n=1425$), (ii) tandem genes that are more than 400-bp apart ($n=855$), (iii) convergent genes that are more than 200-bp apart ($n=452$), (iv) convergent genes that are more than 400-bp apart ($n=208$). Orange rectangles represent open reading frames and black arrows 500-bp of upstream and downstream flanking DNA sequences. (c) Schematic of the *ura4* transgene that was integrated into wild-type and P_{nmt1} -*dis3* strains. The termination element of the *snR99* gene⁶³ was inserted upstream of the *ura4* gene. Arrows

indicate the position of the primer pairs used for RNA and ChIP analyses in (g) and (h), respectively. (d) Northern blot of total RNA prepared from wild-type cells that have a normal *ura4* gene (lanes 1 and 3) or a *ura4* gene in which the *snR99* terminator (*snR99T*) was inserted upstream (lanes 2 and 4), as described in (c). Note that *ura4* mRNAs are not detected in cells that express *ura4* with the *snR99* terminator (lane 2). Read-through *snR99T-ura4* transcripts were detected at low levels on a long exposure in wild-type cells (lane 4). (e) Ten-fold serial dilutions of wild-type cells transformed with the empty vector, the normal *ura4* construct, or the *snR99T-ura4* construct were spotted on uracil-supplemented (left) or uracil-free (right) media. Cells with the *snR99T-ura4* construct do not grow on uracil-free medium. (f) Northern analysis of RNA from the indicated strains that were previously transformed with the *snR99T-ura4* construct and that were grown in the absence (-) or presence (+) of thiamine. The arrow points to the *snR99T-ura4* chimeric RNA, which accumulates in Dis3-depleted cells (lane 4). (g) RT-qPCR analysis of *snR99T-ura4* RNA levels using the strains described in (f). (h) ChIP analysis of RNAPII at the promoter and ORF regions of the *snR99T-ura4* construct. RNAPII density values are relative to cells grown in the absence of thiamine. Note the greater RNAPII density in the ORF region of *ura4* relative to the promoter region, indicative of increased transcription through the *snR99* terminator in Dis3-depleted cells. Error bars, s.d. (n = 3 biological replicates from independent cell cultures). *P* values are from two-tailed Student's *t* test.

The *S. pombe* exosome is present at RNAPII-transcribed genes

To begin to explore the mechanism by which the RNA exosome promotes transcription termination, we first assessed whether subunits of the core exosome complex are associated with sites of RNAPII transcription by ChIP assays, which have previously been used to show the presence of exosome subunits at actively transcribed genes in *S. cerevisiae* (Hieronymus *et al.*, 2004), *Drosophila* (Andrulis *et al.*, 2002), and humans (Basu *et al.*, 2011). Consistent with these observations, ChIP assays using tagged versions of Dis3, Rrp41, Csl4, and Rrp4 showed that these four exosome subunits are associated with sites of snoRNA and mRNA transcription (Fig. 4a-b). Importantly, exosome recruitment to a gene is dependent on active transcription (Fig. 4c-d). Thus, similar to the aforementioned species (Andrulis *et al.*, 2002, Basu *et al.*, 2011, Hieronymus *et al.*, 2004), our data indicate that the fission yeast exosome is present along actively transcribed genes, supporting the involvement of the exosome in transcription termination.

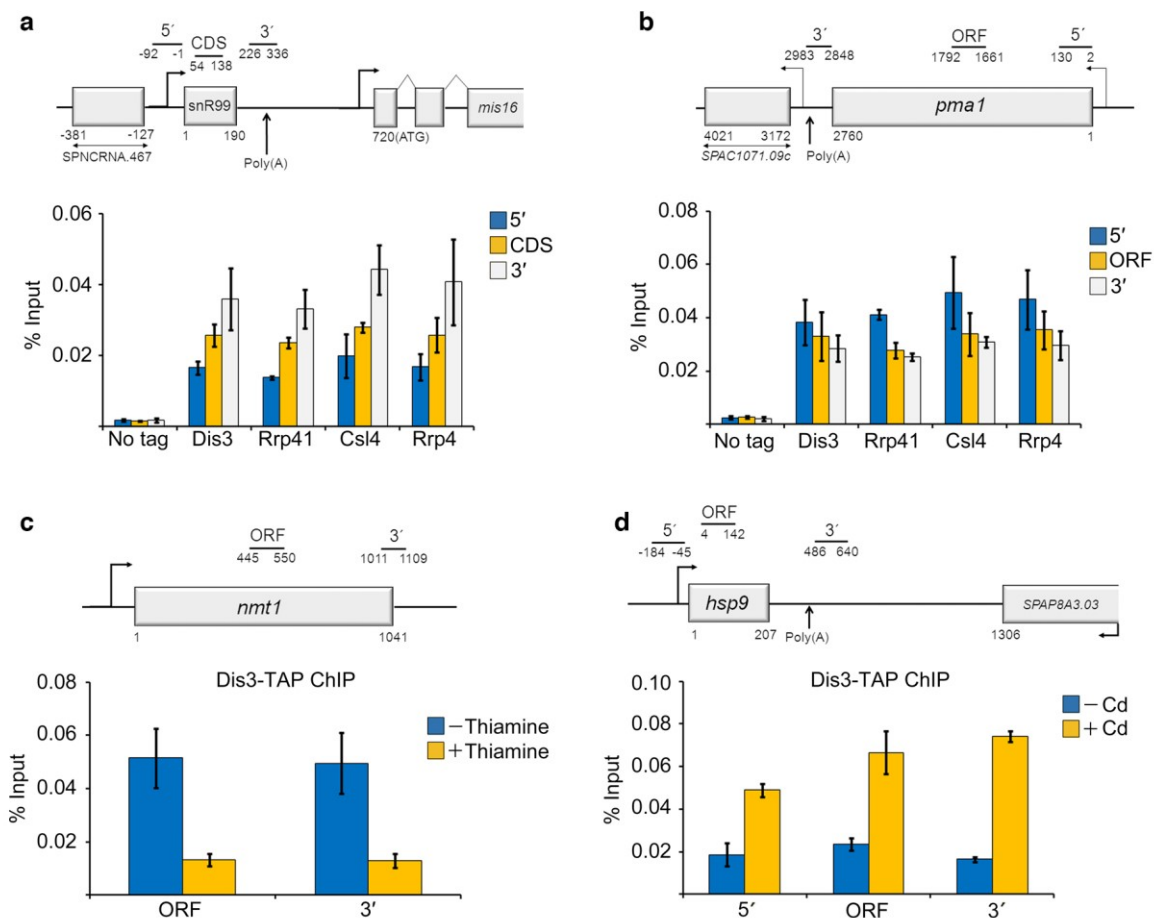


Figure 4. Transcription-dependent recruitment of core exosome subunits to genes.

(a-b) Schematic of *snR99* snoRNA (a) and *pma1* mRNA (b) genes. Bars above the genes show the positions of PCR products used for ChIP analyses. ChIP assays were performed using TAP-tagged versions of Dis3, Rrp41, Csl4, and Rrp4 as well as with an untagged control strain. The input and coprecipitated DNA were quantified by qPCR along the *snR99* (a) and *pma1* (b) genes. (c) ChIP assays using extracts of cells that expressed Dis3-TAP and that were previously grown in the absence (-) and presence (+) of thiamine. The input and coprecipitated DNA were quantified by qPCR along the *nmt1* gene using specific primers, as shown above the graph. Dis3 levels at the *nmt1* gene were reduced in conditions that repress *nmt1* transcription (+ thiamine). (d) ChIP assays using extracts of cells that expressed Dis3-TAP and that were previously grown in the absence (-) and presence (+) of 1 mM cadmium (Cd) for 60 min. The input and coprecipitated DNA were quantified by qPCR along the *hsp9* gene using specific primers, as shown above the graph. Dis3 levels at the *hsp9* gene were increased in conditions that activate *hsp9* expression (+ Cd). Bars indicate either the range of measured values from two independent experiments (d) or the s.d. (a-c) from three experiments using independent cell cultures.

Dis3 3'→5' EXO activity is necessary for RNAPII termination

The single catalytic subunit of the core exosome, Dis3, catalyzes both 3'-5' exonucleolytic (exo) and endonucleolytic (endo) RNA cleavage (Chlebowski *et al.*, 2013, Schneider et Tollervey, 2013). Endonucleolytic cleavage by the RNA exosome could promote transcription termination by a mechanism similar to the termination pathway triggered by the endonucleolytic cleavage catalyzed by the 3' end mRNA processing machinery (Mischo et Proudfoot, 2013). To test this possibility, we generated *dis3* alleles that express single amino acid substitutions at conserved aspartic acid residues (Supplementary Fig. 5a) that were previously shown in *S. cerevisiae* to be devoid of exonucleolytic or endonucleolytic activities (Lebreton *et al.*, 2008). The different *dis3* alleles were chromosomally-integrated as single copy into the P_{nmt1}-*dis3* conditional strain and expressed at levels similar to wild-type Dis3 (Supplementary Fig. 5b). After repression of endogenous *dis3* by thiamine, the expression of wild-type as well as endonucleolytic, exonucleolytic, and double (endo-exo-) mutant versions of *S. pombe* Dis3 resulted in growth phenotypes (Supplementary Fig. 5c-d) similar to the phenotypes of the corresponding *S. cerevisiae* *dis3* mutants (Lebreton *et al.*, 2008, Schaeffer *et al.*, 2009). We next used RNAPII ChIP assays to examine the extent by which the endo- and exo-inactive versions of Dis3 restored the transcription termination defects induced by the depletion of endogenous Dis3. As a control, expression of wild-type Dis3 in the P_{nmt1}-*dis3* strain prevented the thiamine-dependent increase in RNAPII levels downstream of genes (Fig. 5a-d). Strikingly, whereas the endonucleolytically-deficient version of Dis3 fully restored RNAPII density downstream of *snR3* and *pma1* genes, the exonuclease mutant showed a significant increase in RNAPII density at the 3' end of genes (Fig. 5c-d). Consistent with these RNAPII ChIP data, RNA analysis showed the accumulation of 3'-extended transcripts in cells that expressed the exonuclease-deficient version of Dis3, but not in cells expressing the endonucleolytically inactive version (Supplementary Fig. 5e). These results indicate that the 3'-5' exonucleolytic activity, but not the endonucleolytic activity, of Dis3 is required for exosome-dependent transcription termination.

a

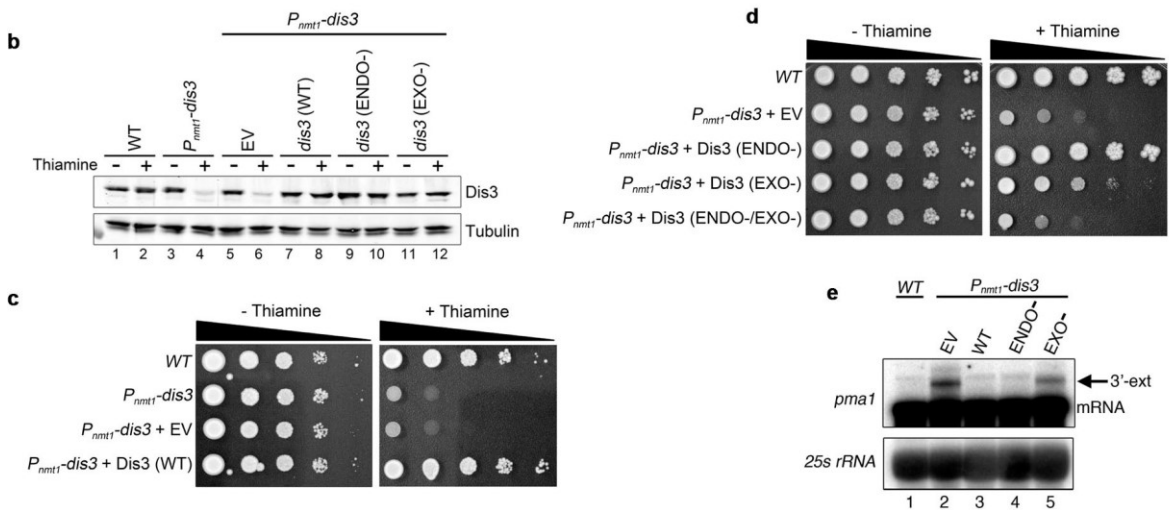
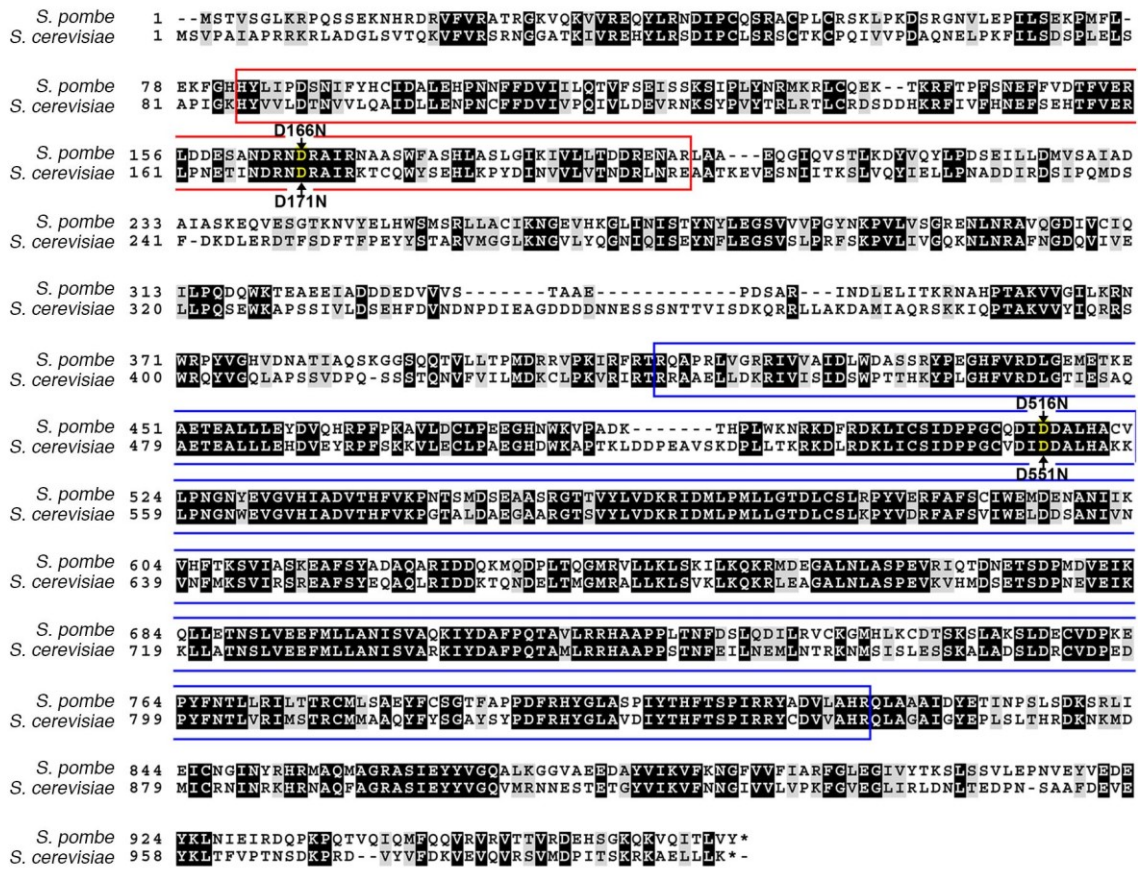


Figure S5. Expression and growth phenotypes of D166N (endo-), D516N (exo-), and D166N D516N (endo- exo-) versions of *S. pombe* Dis3.

(a) Amino acid sequence alignment of *S. pombe* and *S. cerevisiae* Dis3 using Clustal Omega. Identical and similar residues are highlighted in black and gray, respectively. The PIN and RNB domains of *S. pombe* and *S. cerevisiae* Dis3 are boxed in red and blue, respectively. To create an endonucleolytic mutant of *S. pombe* Dis3, aspartic acid (D) residue 166 within the PIN domain was substituted to asparagine (D166N), a substitution homologous to D171N of *S. cerevisiae* Dis3 that results in a version of Dis3 devoid of endonucleolytic activity⁶⁴. To

create an exonucleolytic mutant of *S. pombe* Dis3, aspartic acid (D) residue 516 within the RNB domain was substituted to asparagine (D516N), a substitution homologous to D551N of *S. cerevisiae* Dis3 that results in a version of Dis3 devoid of exonucleolytic activity⁶⁴. (b) Western blot analysis of extracts prepared from wild-type cells (lanes 1-2) as well as from P_{nmt1} -dis3 cells in which an empty vector (lanes 5-6) or constructs that express wild-type Dis3 (lanes 7-8), endo mutant (lanes 9-10), and exo mutant (lanes 11-12) versions of Dis3. Cells grown in the presence (+) of thiamine were depleted for endogenous Dis3 (lanes 4, 6, 8, 10, and 12). (c) Ten-fold serial dilutions of wild-type and P_{nmt1} -dis3 cells as well as P_{nmt1} -dis3 cells that were transformed with an empty vector (EV) or a construct that expresses wild-type Dis3 (WT) were spotted on thiamine-free (left) or thiamine-containing (right) minimal media. (d) Ten-fold serial dilutions of wild-type cells as well as P_{nmt1} -dis3 cells that were transformed with an empty vector (EV) or constructs that express endonuclease-deficient (endo-) Dis3, exonuclease-deficient (exo-) Dis3, and the endo- and exo-deficient double mutant (endo- exo-). Cells were spotted on thiamine-free (left) or thiamine-containing (right) minimal media. The growth phenotypes of the *S. pombe* Dis3 endo-, exo-, and double endo-exo- mutant strains showed the same phenotypes of the corresponding *S. cerevisiae* mutants strains^{64,65}: the endo mutant did not show growth defects, whereas growth of the exo mutant was impaired. Introducing the D166N substitution of the endonuclease domain into the exo mutant (endo- exo- double mutant) exacerbated the growth phenotype of the single exo mutant, consistent with the view that both catalytic activities are required for cell viability^{64,65}. (e) Northern analysis of the *pma1* mRNA using the indicated strains grown in the presence of thiamine. The arrow indicates the read-through *pma1* transcript.

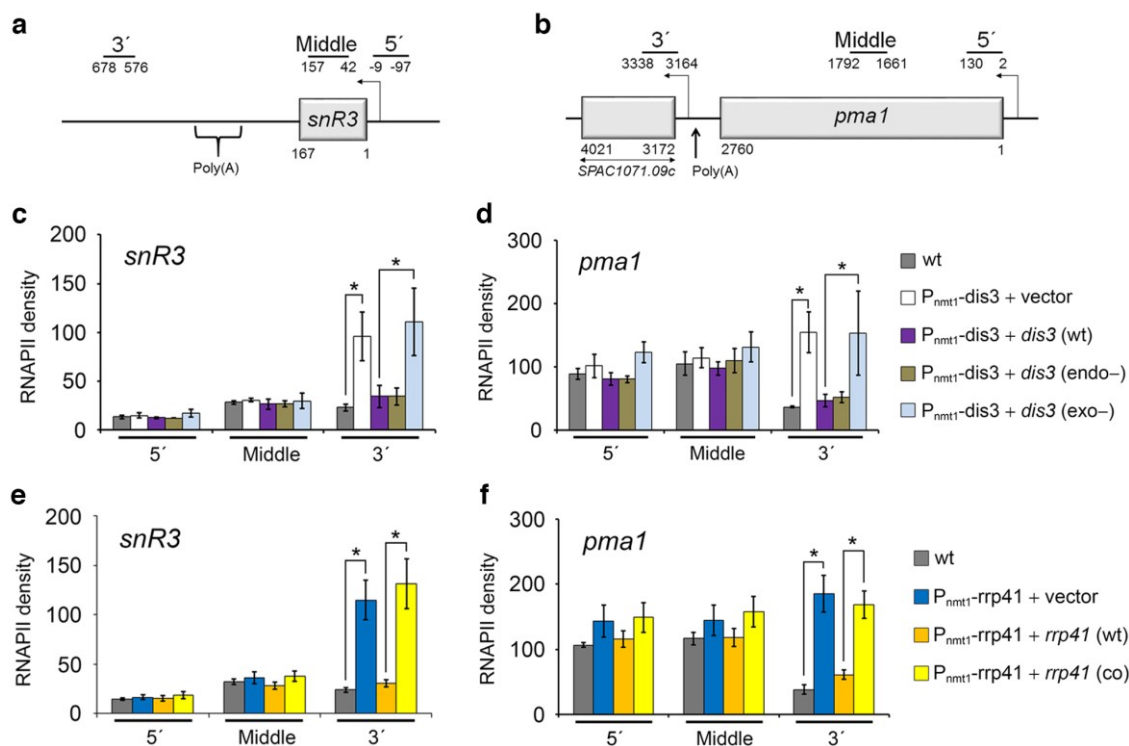


Figure 5. The 3'-5' exonucleolytic activity of Dis3 and the exosome central channel are required for transcription termination.

(a-b) Schematic representation of *snR3* snoRNA gene (a) and *pma1* mRNA (b) genes. Bars above the genes show the positions of PCR products used for ChIP analyses (c-f). (c-d) ChIP analysis of RNAPII density along (5', middle, and 3') the *snR3* (c) and *pma1* (d) genes using extracts prepared from the P_{nmt1} -*dis3* conditional strain that was previously transformed with constructs expressing the wild-type Dis3 (wt), an endonuclease-deficient Dis3 (endo-), a 3'-5' exonuclease-deficient Dis3 (exo-), and the vector control. (e-f) ChIP analysis of RNAPII along the *snR3* (e) and *pma1* (f) genes using extracts prepared from the P_{nmt1} -*rrp41* conditional strain that was previously transformed with constructs expressing the wild-type Rrp41 (wt), the channel-occluded (co) version of Rrp41, and the empty vector control. The wild-type strain was also analyzed as a control in thiamine-supplemented medium. Error bars, s.d. (n = 3 biological replicates from independent cell cultures). * $P < 0.05$ (two-tailed Student's *t* test).

The central channel is required for RNAPII termination

We also examined whether access through the exosome central channel was required for transcription termination by introducing an 11-residue insertion in Rrp41 (Supplementary Fig. 6a) that impedes RNA from threading through the channel (Wasmuth et Lima, 2012). After repression of endogenous Rrp41 by thiamine, expression of the *rrp41* mutant allele into the P_{nmt1} -*rrp41* conditional strain failed to support the essential functions of the RNA exosome (Supplementary Fig. 6b-c), but did not impair exosome assembly (Supplementary Fig. 6d). Importantly, blocking the channel of the RNA exosome resulted in read-through transcription (Fig. 5e-f), indicating that passage through the central channel is required for transcription termination by the core exosome.

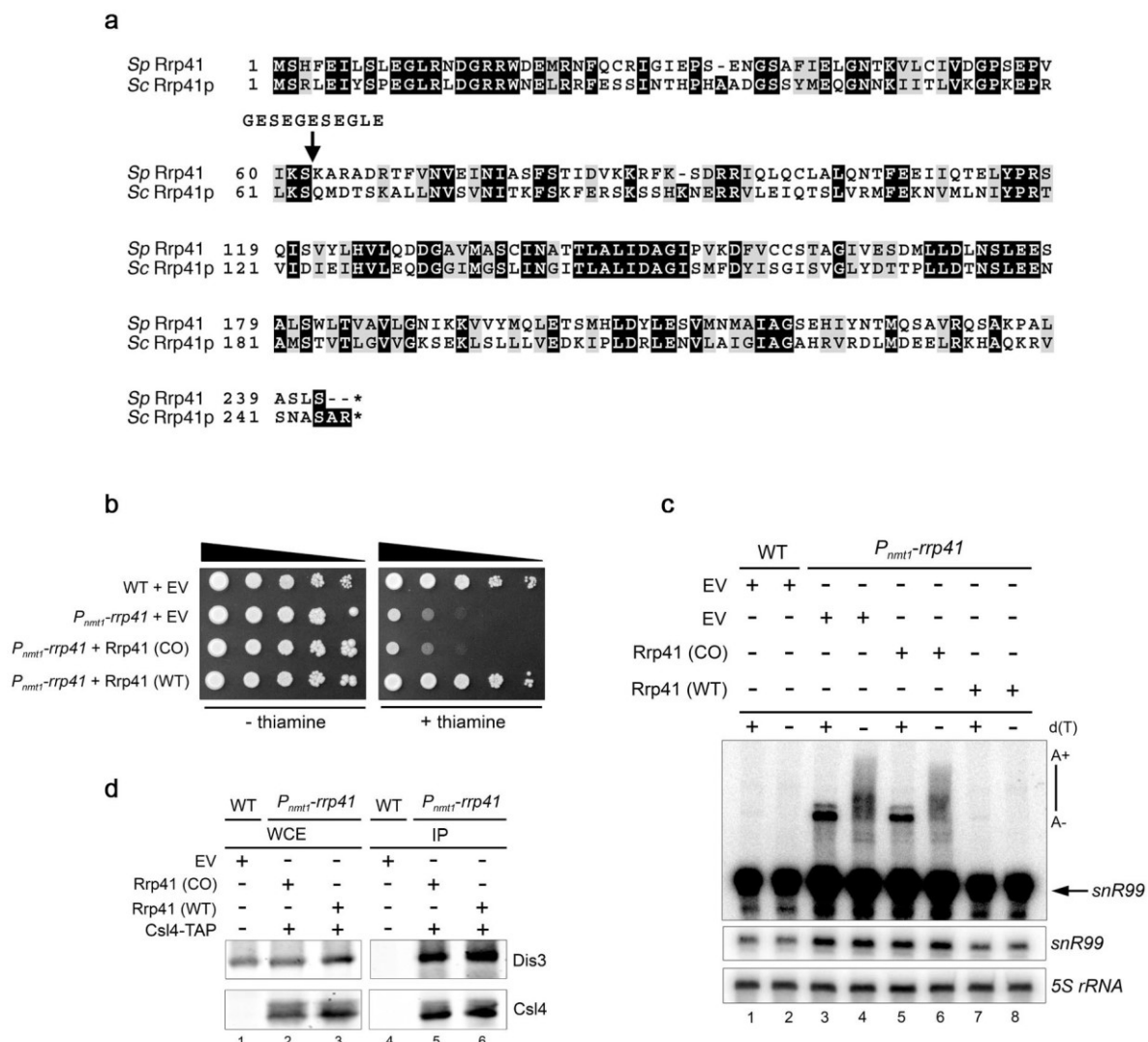


Figure S6. The central channel is essential for viability and snoRNA processing, but does not impair exosome assembly, in *S. pombe*.

(a) Amino acid sequence alignment of *S. pombe* (*Sp*) and *S. cerevisiae* (*Sc*) Rrp41 using Clustal Omega. Identical and similar residues are highlighted in black and gray, respectively. The central channel of the *S. pombe* exosome was occluded by inserting an 11-amino acid sequence (GESEGESEGLE) between serine-62 and lysine-63. This insertion is similar in length and amino acid composition to a previously described insertion that was shown to physically occlude the central channel of the *S. cerevisiae* RNA exosome⁶⁶. (b) Ten-fold serial dilutions of wild-type (WT) and *P_{nmt1}-rrp41* cells that were transformed with an empty vector (EV) or constructs that express wild-type (WT) or channel-occluded (CO) Rrp41. Cells were spotted on thiamine-free or thiamine-containing minimal media. The channel-occluded version of Rrp41 caused growth defect similar to *P_{nmt1}-rrp41* cells transformed with the empty vector. As expected, a wild-type version of Rrp41 restored the growth defect resulting from depletion of endogenous Rrp41. (c) Total RNA prepared from wild-type (lanes 1-2) and *P_{nmt1}-rrp41* (lanes 3-8) strains was treated with RNase H in the presence of a DNA oligonucleotide complementary to H/ACA class snoRNA snR99. RNase H reactions were performed in the presence (+) or absence (-) of oligo(dT). 5S rRNA was used as a loading

control. Expression of a channel-occluded (CO) version of Rrp41 accumulated levels of 3'-extended polyadenylated snR99 precursors (lanes 5-6) similar to P_{nmt1} -rrp41 cells transformed with the empty vector (EV) control (lanes 3-4). (d) A channel-occluded version of Rrp41 does not interfere with core exosome assembly. Immunoblot analysis of whole-cell extracts (WCE; lanes 1-3) and IgG-sepharose precipitates (IP; lanes 4-6) prepared from wild-type (WT) cells transformed with an empty vector (EV) or from P_{nmt1} -rrp41 cells expressing a TAP-tag version of Csl4 transformed with constructs expressing either wild-type (WT; lanes 3 and 6) or channel-occluded (CO; lanes 2 and 5) versions of Rrp41. Cells were grown in the presence of thiamine to repress endogenous Rrp41. Immunoprecipitation of the exosome "cap" protein, Csl4-TAP, in the presence of wild-type (lane 6) or channel-occluded (lane 5) versions of Rrp41 recovered similar levels of Dis3, which is located at the bottom of the core exosome, supporting proper exosome assembly.

Exosome-dependent termination involves RNAPII backtracking

Because transcription termination by the exosome requires a functional exonuclease domain, a free single-stranded 3' end substrate must be available in the context of the transcription process. One transcriptional state in which a free single-stranded RNA 3' end can be accessible occurs on arrested and backtracked RNAPII (Nudler, 2012). During backtracking, the catalytic site of RNAPII becomes disengaged from the RNA 3' end and RNAPII moves backward, causing the 3' end of the nascent RNA to extrude of the polymerase through the secondary channel (Cheung et Cramer, 2011). Backtracked RNAPII can be rescued by a slow intrinsic hydrolyzing activity of RNAPII that is stimulated by the transcription factor TFIIIS (Izban et Luse, 1992) (Tfs1 in *S. pombe*), which cleaves the extruded RNA and causes the re-alignment of the RNA 3' end in the catalytic site, allowing transcription to resume. We thus hypothesized that exosome-mediated transcription termination depends on RNAPII backtracking, whereby extrusion of the nascent RNA provides a substrate for the core exosome, triggering polymerase release together with the concomitant degradation of the read-through RNA (Fig. 6a). In exosome-depleted cells (Fig. 6b), TFIIIS would outcompete exosome-mediated termination, promoting the reactivation of backtracked RNAPII complexes until they reach distal contexts for 3' end processing and polyadenylation, resulting in the production of read-through transcripts.

To test this model, we deleted *tfs1* (TFIIS) in wild-type and P_{nmt1} -dis3 cells. The absence of Tfs1 in exosome-depleted cells resulted in reduced levels of read-through RNAs relative to the single exosome mutant (Fig. 6c, compare lanes 2 and 4). In contrast, levels of regular length mRNAs were not affected (Fig. 6c), indicating that the absence of Tfs1 specifically reduced the levels of read-through transcripts relative to regular mRNAs in Dis3-depleted cells (Fig. 6d). To support these RNA analyses, we examined RNAPII profiles by ChIP-seq. The absence of TFIIS in *S. pombe* did not impair RNAPII occupancy in the coding regions (Fig. 6e, compare red and black profiles upstream of poly(A) site). This observation is consistent with the absence of a detectable decrease in mRNA levels in *tfs1Δ* cells (Fig. 6c) and data using TFIIS mutants in *S. cerevisiae* (Mason et Struhl, 2005, Sigurdsson *et al.*, 2010). In contrast to coding regions, the absence of Tfs1 resulted in premature dissociation of RNAPII downstream of poly(A) sites (Fig. 6e, compare red and black profiles downstream of poly(A) site). Importantly, premature termination in the *tfs1Δ* mutant required the RNA exosome, as depletion of Dis3 suppressed the earlier termination phenotype seen in the *tfs1Δ* single mutant (Fig. 6e, compare red and purple profiles). Likewise, the absence of Tfs1 suppressed read-through transcription in Dis3-depleted cells (Fig. 6e, compare blue and purple profiles). These data strongly suggest that the RNA exosome and TFIIS compete for backtracked RNAPII at the 3' end of genes.

Even in the absence of TFIIS, arrested and backtracked RNAPII can be reactivated by the intrinsic RNA cleavage activity of the polymerase large subunit, albeit less efficiently (Sigurdsson *et al.*, 2010). We thus tested the effect of a previously described mutant of TFIIS that inhibits the cleavage activity of RNAPII and prevents the rescue of backtracked transcription complexes (Sigurdsson *et al.*, 2010) (Supplementary Fig. 7a-b). Notably, expression of this dominant-negative TFIIS mutant almost completely abolished the production of exosome-dependent read-through transcripts (Fig. 6f, compare lane 4 to lanes 1-2; quantification in Supplementary Fig. 7c), indicating that the remaining read-through RNAs produced in the absence of Tfs1 (Fig. 6c, lane 4 and Fig. 6f, lane 3) result from the non-stimulated intrinsic cleavage activity of RNAPII. As a control, expression of wild-type Tfs1 in the P_{nmt1} -dis3 *tfs1Δ* double mutant strain restored the levels of read-through RNA detected in the single P_{nmt1} -dis3 mutant (Fig. 6f, compare lane 5 to lane 2). Together,

these results indicate that the detection of 3'-extended RNAs and read-through RNAPII in exosome-depleted cells depend on TFIIS. Given the established role of TFIIS in the reactivation of backtracked RNAPII *in vitro* and *in vivo* (Cheung et Cramer, 2011, Churchman et Weissman, 2011, Sigurdsson *et al.*, 2010), we conclude that exosome-mediated transcription termination involves RNAPII backtracking.

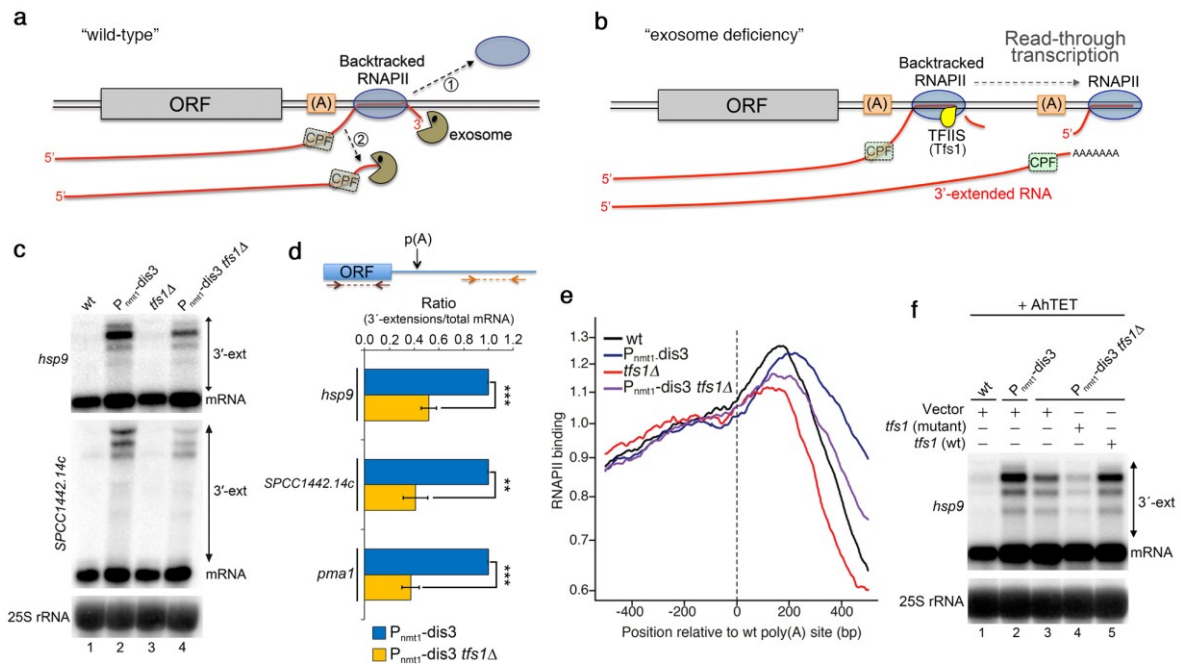


Figure 6. Transcription termination by the exosome is mechanistically linked to RNAPII backtracking.

(a) Model in which backtracking provides a free RNA 3' end for the core exosome, promoting termination (1) and degradation (2) of potential read-through RNAs uncleaved by the Cleavage and Polyadenylation Factor (CPF) machinery. (b) In exosome mutants, TFIIS (Tfs1 in *S. pombe*)-stimulated RNAPII-mediated RNA cleavage outcompetes termination of backtracked RNAPII by the exosome, allowing transcription to resume until 3' end processing at distal cleavage sites. (c) Northern analysis of RNA from the indicated strains grown in the presence of thiamine. (d) Ratio of 3'-extended transcripts relative to total mRNA as measured by real-time RT-PCR using primers located downstream of poly(A) sites and in ORF sequences. Data in the $P_{nmt1}\text{-dis3 } tfs1\Delta$ double mutant are relative to the single $P_{nmt1}\text{-dis3}$ mutant. Error bars, s.d. ($n = 3$ biological replicates from independent cell cultures). ** $p < 0.01$ and *** $p < 0.001$ (two-tailed Student's *t* test). (e) Cumulative RNAPII profile in the indicated strains. Curves show the sum of normalized ChIP-seq sequencing scores over a genomic region covering the main poly(A) site. (f) Northern analysis of RNA from wild-type (lane 1), $P_{nmt1}\text{-dis3}$ (lane 2), and $P_{nmt1}\text{-dis3 } tfs1\Delta$ (lanes 3-5) cells transformed with a vector that expresses wild-type (lane 5) and mutant (lane 4) Tfs1 under the control of tetracycline

(AhTET), as well as the empty vector (lanes 1-3). Cells were grown for 6h in the presence of thiamine, after which AhTET was added for another 6h.

Backtracking is a major mechanism underlying pausing of RNA polymerases (Churchman et Weissman, 2011). We therefore tested the impact of mutations at the 3' end of the *ura4* gene (Supplementary Fig. 7d) that perturb RNAPII pausing (Birse *et al.*, 1997) on exosome-dependent termination. Consistent with previous findings (Birse *et al.*, 1997), removal of these transcriptional pause signals resulted in transcription termination defects, as measured by the 4-fold increase in RNAPII density at the 3' end of the mutant *ura4* gene relative to wild-type *ura4* (Supplementary Fig. 7e). Next, wild-type and mutant *ura4* constructs were chromosomally integrated into the P_{nmt1} -dis3 conditional strain, and RNAPII density was measured at the 3' end of *ura4* by ChIP assays using cells cultured in the presence and absence of thiamine. Consistent with our previous results, we found a thiamine-dependent increase in RNAPII density at the 3' end of the wild-type *ura4* gene in P_{nmt1} -dis3 cells relative to control cells (Supplementary Fig. 7f; see WT *ura4* 3'). In contrast, read-through polymerases that result from loss of exosome function were no longer detected at the 3' end of the pause-defective *ura4* mutant (Supplementary Fig. 7f; see Mutant *ura4* 3'). Together, these results indicate that transcription termination by the RNA exosome is mechanistically linked to RNAPII pausing and backtracking.

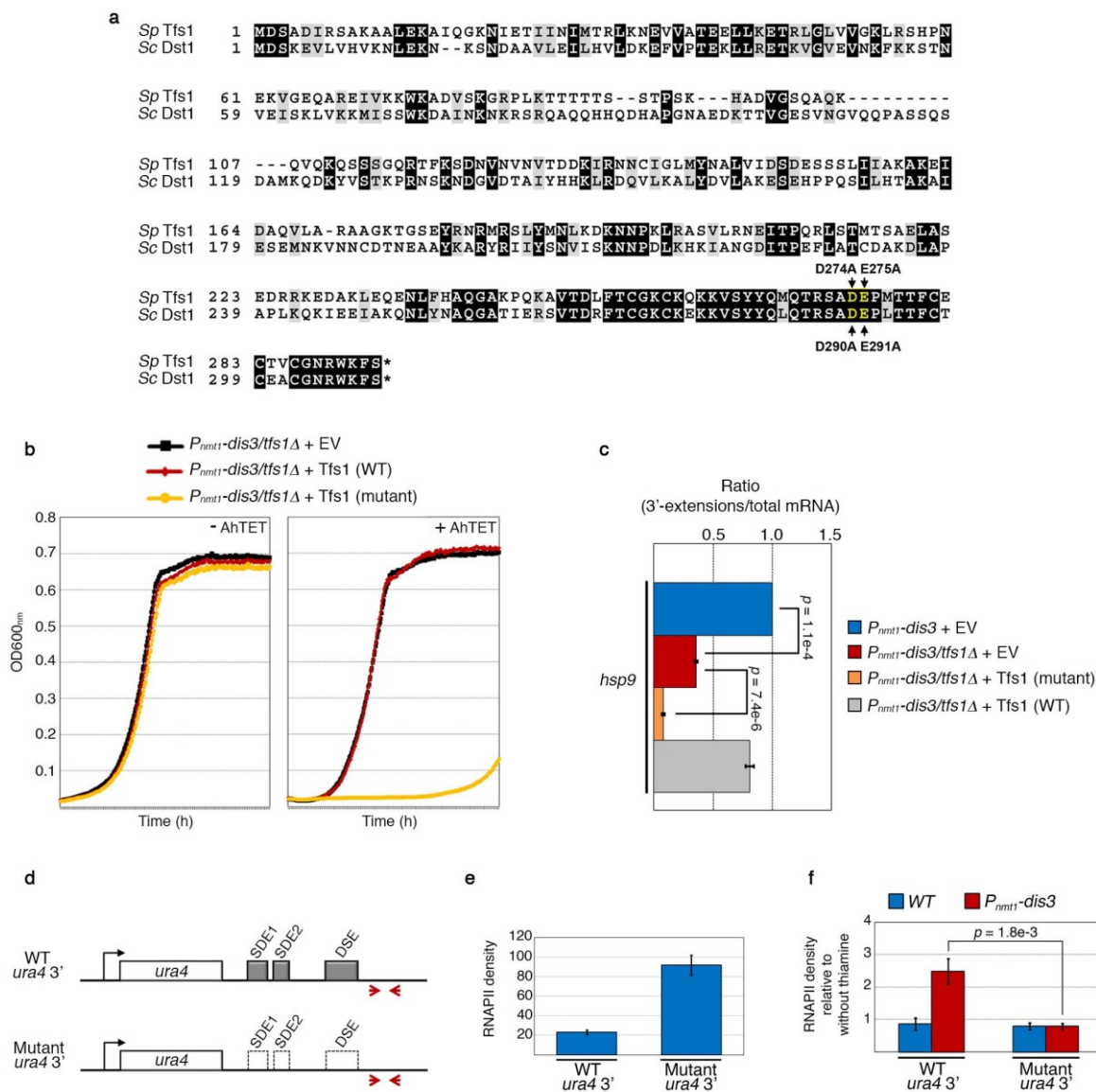


Figure S7. Transcription termination by the RNA exosome is mechanistically linked to RNAPII backtracking and pausing.

(a-c) A Tfs1 mutant that lacks transcript cleavage stimulatory activity and inhibits unstimulated cleavage results in a more pronounced suppression of read-through RNAs relative to the *tfs1* deletion in Dis3-depleted cells. (a) The *S. pombe* transcription elongation factor Tfs1 possesses a highly conserved acidic loop important for cell viability and transcript cleavage. Sequence alignments of the transcription elongation factors Tfs1 of *S. pombe* and Dst1p (TFIIS) of *S. cerevisiae*. Arrows indicate the acidic amino acids involved in stimulating transcript cleavage inside the catalytic center of RNAPII. In order to abolish transcript cleavage, the conserved Asp²⁷⁴-Glu²⁷⁵ dipeptide of the acidic loop were changed to alanine residues, as described previously in *S. cerevisiae*⁶⁷. (b) Analysis of cell growth by optical density measurements (OD₆₀₀). *P_{nmt1}-dis3 tfs1Δ* cells that were previously transformed with DNA constructs expressing wild-type or mutant (D274A E275A) versions of Tfs1 under the control of a tetracycline-inducible promoter, as well as the empty vector (EV) control, were grown in media containing 7.5 μM of the non-metabolic inducer

anhydrotetracycline hydrochloride (+AhTET; right) or in mock-treated media (-AhTET; left). Samples were prepared in triplicate and cell growth was monitored every 10 min for 48 h. Inducing the expression of Tfs1 D274A E275A (*tfs1* mutant) causes severe growth defect, as shown previously in *S. cerevisiae*⁶⁷. (c) real-time RT-qPCR analysis of *hsp9* read-through and total mRNA using RNA obtained from cell grown as described in Fig. 6f. The *hsp9* 3'-extensions/total RNA ratio was arbitrarily set to 1.0 for the P_{nmt1}-dis3 + EV strain. Error bars, s.d. (n = 3 biological replicates from independent cell cultures). *P* values are from two-tailed Student's *t* test. (d-f) Termination of transcription by the exosome requires RNA polymerase II pausing. (d) Schematic of wild-type and mutant *ura4* constructs used in (e) and (f). Site-determining elements (SDE) 1-2 and downstream sequence elements (DSE) important for RNAPII pausing⁶⁸ are shown. The red arrows indicate the position of the PCR primers used to analyze RNAPII density at the 3' end of *ura4*. (e) ChIP analysis of RNAPII at the 3' end of the wild-type and mutant *ura4* gene in wild-type cells. Disruption of transcriptional pause signals (mutant) resulted in transcription termination defects, as measured by the 4-fold increase in RNAPII density at the 3' end of the mutant *ura4* gene relative to wild-type *ura4*, consistent with previous findings⁶⁸. (f) Wild-type and mutant *ura4* constructs were inserted into the P_{nmt1}-dis3 conditional strain, and RNAPII density was measured at the 3' end of *ura4* by ChIP assays using cells cultured in the presence and absence of thiamine. A thiamine-dependent increase in RNAPII density was detected at the 3' end of the wild-type *ura4* gene (WT *ura4* 3') in P_{nmt1}-dis3 cells (red) relative to wild-type cells (blue), consistent with termination defects in Dis3-depleted cells. In contrast, read-through polymerases that result from loss of exosome function (red) were no longer detected at the 3' end of the pause-defective *ura4* mutant (mutant *ura4* 3'). RNAPII density values are relative to cells grown in the absence of thiamine. (e-f) Error bars, s.d. (n = 3 biological replicates from independent cell cultures). *P* value is from two-tailed Student's *t* test.

DISCUSSION

In this study, we identified an unsuspected role for the RNA exosome in a transcription termination pathway coupled to RNAPII pausing and arrest. We provide evidence into how this termination-coupled RNA degradation pathway promotes the removal of backtracked RNAPII, thereby preventing the production of read-through transcripts and transcription interference into neighboring genes. These findings provide mechanistic insights into the crosstalk between RNA surveillance, 3' end processing, and transcription termination.

Termination of backtracked RNAPII by the RNA exosome

Genome-wide studies indicate that transcription elongation by RNAPII is not a uniform process(Gomez-Herreros *et al.*, 2012), but that RNAPII frequently pauses and backtracks after facing obstacles in the template DNA, such as nucleosomes(Churchman et Weissman, 2011). Reactivation of backtracked polymerase by TFIIS-stimulated RNA cleavage is a mechanism known to resume transcription elongation by RNAPII(Cheung et Cramer, 2011, Gomez-Herreros *et al.*, 2012). In principle, a backtracked RNAPII complex could also be removed by disassembling the arrested polymerase, resulting in termination. However, although *in vitro* work has underscored the importance of backtracking in transcription termination by eukaryotic RNA polymerase III(Nielsen *et al.*, 2013) and the prokaryotic RNA polymerase(Gusarov et Nudler, 2001), experimental evidence associating polymerase backtracking and transcription termination has not been disclosed for RNAPII. Our results now provide evidence that the RNA exosome can terminate transcription events associated with backtracked RNAPII. Although the nature of the obstructions promoting RNAPII backtracking at the end of genes remains to be determined, pausing of RNAPII downstream of poly(A) signals is supported by several studies(Anamika *et al.*, 2012, Coudreuse *et al.*, 2010, Glover-Cutter *et al.*, 2008, Gromak *et al.*, 2006), and is consistent with the accumulation of RNAPII beyond annotated poly(A) sites in our ChIP-seq analyses (Fig. 3, Supplementary Fig. 4, and Fig. 6). Backtracking at pause sites can extend to dozens of nucleotides in a biological context(Churchman et Weissman, 2011), a length sufficient to extrude RNA from the funnel-shaped pore of RNAPII(Cheung et Cramer, 2011), exposing an unprotected single-stranded RNA 3' end. Threading of such single-stranded RNA by the unwinding substrate entry pore of the core exosome(Makino *et al.*, 2013) will reach the Dis3 active site for processive 3'-5' exonucleolytic activity, which may displace the paused and less stable polymerase complex. Alternatively, the processive exonucleolytic activity of the exosome may promote further backtracking on reaching the polymerase, which could trigger hairpin-dependent termination(Nielsen *et al.*, 2013). Thus, in addition to proteasome-dependent degradation of RNAPII(Somesh *et al.*, 2005), exosome-dependent termination provides an additional mean to control transcriptional arrest when reactivation mechanisms are not desired.

RNA surveillance by the exosome via termination-coupled RNA decay

Accumulating evidence support that poly(A)-dependent 3' end processing is coupled to transcription termination. Nevertheless, a substantial number of RNAPII fail to terminate upon reaching a poly(A) signal (Mischo et Proudfoot, 2013), presumably due to non-productive endonucleolytic cleavage by the 3' end processing machinery. Although such read-through transcription is known to be coupled to exosome-mediated RNA degradation (Torchet *et al.*, 2002, West *et al.*, 2006), it was unclear how the RNA exosome post-transcriptionally distinguished read-through transcripts from normal mRNAs. The results presented in this study support the idea that termination-coupled RNA degradation by the exosome competes with 3' end processing when RNAPII pauses and backtracks downstream of poly(A) signals (Fig. 7a). In productive cases, 3' end cleavage will occur and the exosome-dependent termination pathway uncovered in this study will not have access to the pre-mRNA, as the pre-mRNA is physically separated from the arrested transcription complex due to 3' end cleavage and polyadenylation (Fig. 7b). Furthermore, in the case of productive 3' end processing, it is tempting to speculate on a double “torpedo” model of transcription termination, whereby the 3'-5' exonucleolytic activity of the core exosome might cooperate with the 5'-3' exonuclease activity of the Rat1 (Xrn2) complex to disengage RNAPII from the DNA template (Fig. 7b). In contrast, in the cases in which 3' end processing does not result in productive endonucleolytic cleavage, the core exosome can promote transcription termination and degrade the uncleaved “aberrant” pre-mRNA (Fig. 6a). By competing with mRNA 3' end processing, transcription termination by the RNA exosome provides a failsafe mechanism that promotes the removal of paused and backtracked RNAPII associated with uncleaved transcripts (Fig. 6a), thereby preventing the potential for read-through transcription via reactivation of arrested RNAPII (Fig. 6b).

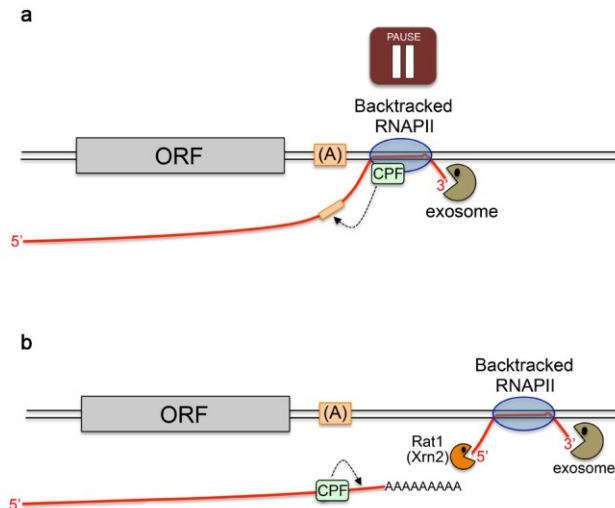


Figure 7. Model of co-transcriptional RNA surveillance by the RNA exosome by competition with 3' end processing. (a) Transcriptional pausing, which is frequently associated with RNAPII backtracking (Churchman et Weissman, 2011, Nudler, 2012), enhances 3' end processing by the Cleavage and Polyadenylation Factor (CPF) complex by facilitating the recognition of poly(A) signals on the nascent transcripts. RNAPII backtracking also provides a free RNA 3' end for exosome-dependent termination, thereby challenging the CPF complex to efficiently process the nascent pre-mRNA. In the cases of unproductive cleavage by the CPF, exosome-dependent termination will promote the removal of backtracked RNAPII together with the degradation of the uncleaved transcripts (Fig. 6a), thereby preventing the potential for read-through transcription. (b) Productive 3' end processing by the CPF releases the pre-mRNA for polyadenylation and provides a substrate for the 5'-3' exonuclease Rat1 (Xrn2) to promote termination of transcription.

In summary, our findings provide evidence for a new mechanism of transcription termination coupled to RNAPII backtracking that functions in co-transcriptional RNA surveillance by the nuclear exosome. Because backtracked RNAPII was a proposed entry point for the human exosome during class switch recombination at transcribed immunoglobulin loci (Basu *et al.*, 2011), we predict that the mechanism described in this study highlights a general feature of co-transcriptional nuclear surveillance in eukaryotes.

ACKNOWLEDGMENTS

We thank M. Bühler and P-E. Jacques for helpful discussions; M. Yamamoto (University of Tokyo) and D. Hermand (University of Namur) for strains; M. Durant and E. Lapointe for ChIP-seq libraries; the sequencing platform of the McGill University and Génome Québec Innovation Centre; S. Abou Elela, E. Lei, and A. Morillon for critical reading of the manuscript. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) to F.B., a Wellcome Trust Senior Investigator Award to J.B., and the Medical Research Council to S.M.. J-F.L. was supported by an Alexander Graham-Bell Doctoral Scholarship from NSERC. F.B. is a Canada Research Chair in Quality Control of Gene Expression.

AUTHOR CONTRIBUTIONS

J-F. L., M.L., and F.B. conceived and designed the research project; J-F. L. and M.L. performed the experiments; S.M. and S.A. prepared the RNA-seq libraries in collaboration with J.B.; S.M. performed computational analyses of RNA-seq and ChIP-seq experiments; J-F. L., M.L., S.M., and F.B. analyzed the data; F.B. wrote the manuscript; all authors discussed the results and commented on the manuscript.

REFERENCES

- 1 Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* 15, 163-175 (2014).
- 2 Mischo, H. E. & Proudfoot, N. J. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta* 1829, 174-185 (2013).
- 3 Dengl, S. & Cramer, P. Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* 284, 21270-21279 (2009).

- 4 Jimeno-Gonzalez, S., Schmid, M., Malagon, F., Haaning, L. L. & Jensen, T. H. Rat1p maintains RNA polymerase II CTD phosphorylation balance. *RNA* 20, 551-558 (2014).
- 5 Kim, M. et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517-522 (2004).
- 6 Luo, W., Johnson, A. W. & Bentley, D. L. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* 20, 954-965 (2006).
- 7 Pearson, E. L. & Moore, C. L. Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor Rat1. *J Biol Chem* 288, 19750-19759 (2013).
- 8 West, S., Gromak, N. & Proudfoot, N. J. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 522-525 (2004).
- 9 Anamika, K., Gyenis, A., Poidevin, L., Poch, O. & Tora, L. RNA polymerase II pausing downstream of core histone genes is different from genes producing polyadenylated transcripts. *PLoS One* 7, e38769 (2012).
- 10 Coudreuse, D. et al. A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in *S. pombe*. *Curr Biol* 20, 1053-1064 (2010).
- 11 Glover-Cutter, K., Kim, S., Espinosa, J. & Bentley, D. L. RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nat Struct Mol Biol* 15, 71-78 (2008).
- 12 Gromak, N., West, S. & Proudfoot, N. J. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol* 26, 3986-3996 (2006).
- 13 Marguerat, S. et al. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* 151, 671-683 (2012).
- 14 Chlebowski, A., Lubas, M., Jensen, T. H. & Dziembowski, A. RNA decay machines: the exosome. *Biochim Biophys Acta* 1829, 552-560 (2013).
- 15 Schneider, C. & Tollervey, D. Threading the barrel of the RNA exosome. *Trends Biochem Sci* 38, 485-493 (2013).

- 16 Januszyk, K. & Lima, C. D. The eukaryotic RNA exosome. *Curr Opin Struct Biol* 24, 132-140 (2014).
- 17 Schmid, M. & Jensen, T. H. Transcription-associated quality control of mRNP. *Biochim Biophys Acta* 1829, 158-168 (2013).
- 18 Andrulis, E. D. et al. The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature* 420, 837-841 (2002).
- 19 Basu, U. et al. The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 144, 353-363 (2011).
- 20 Hessle, V. et al. The exosome associates cotranscriptionally with the nascent pre-mRNP through interactions with heterogeneous nuclear ribonucleoproteins. *Mol Biol Cell* 20, 3459-3470 (2009).
- 21 Hieronymus, H., Yu, M. C. & Silver, P. A. Genome-wide mRNA surveillance is coupled to mRNA export. *Genes Dev* 18, 2652-2662 (2004).
- 22 Lim, S. J., Boyle, P. J., Chinen, M., Dale, R. K. & Lei, E. P. Genome-wide localization of exosome components to active promoters and chromatin insulators in *Drosophila*. *Nucleic Acids Res* 41, 2963-2980 (2013).
- 23 Castelnovo, M. et al. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* 20 (2013).
- 24 Shah, S., Wittmann, S., Kilchert, C. & Vasiljeva, L. lncRNA recruits RNAi and the exosome to dynamically regulate *pho1* expression in response to phosphate levels in fission yeast. *Genes Dev* 28, 231-244 (2014).
- 25 Wagschal, A. et al. Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* 150, 1147-1157 (2012).
- 26 Yamanaka, S., Yamashita, A., Harigaya, Y., Iwata, R. & Yamamoto, M. Importance of polyadenylation in the selective elimination of meiotic mRNAs in growing *S. pombe* cells. *EMBO J* 29, 2173-2181 (2010).
- 27 Keller, C., Woolcock, K., Hess, D. & Buhler, M. Proteomic and functional analysis of the noncanonical poly(A) polymerase Cid14. *RNA* 16 (2010).
- 28 Vasiljeva, L. & Buratowski, S. Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell* 21, 239-248 (2006).

- 29 West, S., Gromak, N., Norbury, C. J. & Proudfoot, N. J. Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Mol Cell* 21, 437-443 (2006).
- 30 Lebreton, A., Tomecki, R., Dziembowski, A. & Seraphin, B. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* 456, 993-996 (2008).
- 31 Schaeffer, D. et al. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat Struct Mol Biol* 16, 56-62 (2009).
- 32 Wasmuth, E. V. & Lima, C. D. Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. *Mol Cell* 48, 133-144 (2012).
- 33 Nudler, E. RNA polymerase backtracking in gene regulation and genome instability. *Cell* 149, 1438-1445 (2012).
- 34 Cheung, A. C. & Cramer, P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* 471, 249-253 (2011).
- 35 Izban, M. G. & Luse, D. S. The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'→5' direction in the presence of elongation factor SII. *Genes Dev* 6, 1342-1356 (1992).
- 36 Mason, P. B. & Struhl, K. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell* 17, 831-840 (2005).
- 37 Sigurdsson, S., Dirac-Svejstrup, A. B. & Svejstrup, J. Q. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell* 38, 202-210 (2010).
- 38 Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368-373 (2011).
- 39 Birse, C. E., Lee, B. A., Hansen, K. & Proudfoot, N. J. Transcriptional termination signals for RNA polymerase II in fission yeast. *EMBO J* 16, 3633-3643 (1997).
- 40 Gomez-Herreros, F. et al. One step back before moving forward: regulation of transcription elongation by arrest and backtracking. *FEBS Lett* 586, 2820-2825 (2012).

- 41 Nielsen, S., Yuzenkova, Y. & Zenkin, N. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* 340, 1577-1580 (2013).
- 42 Gusarov, I. & Nudler, E. Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell* 107, 437-449 (2001).
- 43 Makino, D. L., Baumgartner, M. & Conti, E. Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* 495, 70-75 (2013).
- 44 Somesh, B. P. et al. Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell* 121, 913-923 (2005).
- 45 Torchet, C. et al. Processing of 3'-extended read-through transcripts by the exosome can generate functional mRNAs. *Mol Cell* 9, 1285-1296 (2002).

EXPERIMENTAL PROCEDURES

Yeast strains, plasmids and media

A list of all *S. pombe* strains and plasmids used in this study is provided in Supplementary Table 1 & 2. Cells were routinely grown at 30°C in yeast extract medium with amino acid supplements (YES) or Edinburgh minimal media (EMM) supplemented with appropriate amino acids. Conditional strains in which the genomic copy of *dis3*, *rrp41*, and *mtr4* is expressed from the thiamine-repressible *nmt1* promoter (P_{nmt1}) were generated as described previously (Lemieux *et al.*, 2011). P_{nmt1} -dependent expression was repressed by the addition of 60 μ M thiamine to the growth medium for 12-15 hours. For induction of *tfs1* expression and its derivative mutated version, an initial culture grown for 6 hours in presence of thiamine was split into two cultures supplemented with either 0.3% EtOH or 7.5 μ M of anhydrotetracycline hydrochloride (AhTET; Sigma-Aldrich, 94664), and allowed to grow for another 6 hours. Gene disruptions and C-terminal tagging of proteins of interest were performed by PCR-mediated gene targeting (Bahler *et al.*, 1998), using the lithium acetate method for cell transformation. Knockouts strains were confirmed by RT-PCR while

promoter swaps and protein tagging were validated by PCR screening and/or Western blotting. Primer sequences used throughout this study are available upon request.

***ura4* reporter system for transcription termination defects**

Reporter constructs for *snR99* terminator activity were created using the *ade6* integration vector(Lemieux *et al.*, 2011), pFB366 as the host vector and contain 1 kb of *snR99* promoter sequences followed by the *ura4* coding region and an additional 0.5kb of *ura4* 3' UTR sequences, resulting in pFB741. To address the impact of depleting Dis3 on transcription termination, a 332-nt region from the *snR99* terminator sequence(Lemay *et al.*, 2010) was cloned upstream of the *ura4* coding region of pFB741, generating pFB745. Both constructs were confirmed by sequencing. For single integration into the *ade6* locus, pFB741 and pFB745 were linearized and transformed into FBY106 and FBY721 strains. Positive integrants were confirmed by PCR screening using primers located within the *ura4* coding region.

***ura4* construct with alteration in transcriptional pausing elements**

The *ura4* mutant defective in transcriptional pausing (*ura4* 3' mutant) was generated by a multi-step procedure beginning with the PCR-mediated deletion of the DSE element(Birse *et al.*, 1997) from a genomic DNA fragment covering the *ura4* coding region and 0.5kb of 3' intergenic region using a 106-nt-long reverse primer that included 5' and 3' flanking DNA sequences of the 131-nt-long DSE element. The PCR product corresponding to DSE-deleted *ura4* was cloned downstream of *snR99* promoter sequences, generating pFB836. Deletion of SDE1 and mutations in SDE2 were as described previously(Birse *et al.*, 1997) using PCR-based mutagenesis on pFB836, resulting in pFB881. All constructs were confirmed by DNA sequencing. For single integration into the *ade6* locus, pFB741 (*ura4* 3' wt) & pFB881 (*ura4* 3' mutant) were linearized and transformed into FBY106 and FBY721 strains. Positive integrants were confirmed by PCR screening using primers located within the *ura4* coding region.

Tfs1 expression constructs

To generate the Tfs1 expression constructs, the *tfs1* coding region was PCR-amplified from cDNA and cloned into the tetracycline-inducible expression vector pMLtet^{ON} (a generous gift of Martin Luetzelberger) (pFB789), resulting in pFB811. Plasmid pMLtet^{ON} was constructed by replacement of the *nmt1* promoter in pML11 (see below, located between *PstI* and *AgeI*) with a DNA fragment carrying the CaMV35Stet^{ON} promoter and *tet* repressor, which was PCR amplified from pDUAL-tetR-rpsL-neo(Erler *et al.*, 2006). In order to obtain unique restriction sites within the MCS of pMLtet^{ON}, two restriction sites (*BamHI* and *SphI*), were subsequently removed from the pDUAL fragment by site-directed mutagenesis. Plasmid pML11 was constructed by insertion of a new MCS (*PstI-NdeI-AgeI-SphI-BamHI-SalI-SpeI-MluI-SmaI-SacI*) into the *PstI-SacI* site of pJK148. Subsequently, the *nmt1* promoter and -terminator sequences of pREP2(Maundrell, 1993) were ligated into the *PstI-NdeI* and *SmaI-SacI* sites of the new MCS, respectively. The mutated version of *tfs1* was created by site-directed mutagenesis using pFB811 as template changing aspartate-274 and glutamate-275 to alanine residues. All constructs were confirmed by DNA sequencing.

Dis3 expression constructs

To generate the Dis3 expression constructs, the coding region of *dis3* with an additional 0.5kb of promoter and terminator sequences was PCR-amplified from genomic DNA and cloned into pFB366, resulting in pFB693. Endoribonucleolytic (endo-) and exoribonucleolytic (exo-) mutant versions of Dis3 were generated by site-directed mutagenesis using pFB693 as template changing aspartate-166 and aspartate-516 to asparagine residues, resulting in pFB702 (endo-) and pFB707 (exo-), respectively. A catalytically inactive version of Dis3 impaired in both enzymatic activities (endo- exo-) was also generated (pFB708). All constructs were confirmed by sequencing. For single integration into the *ade6* locus, pFB366, pFB693, pFB702, pFB707 and pFB708 were

linearized and transformed into the conditional P_{nmt1} -dis3 strain (FBY721). Positive integrants were confirmed by growth selection on EMM agar plate lacking adenine.

Rrp41 expression constructs

To generate the wild-type (WT) Rrp41 expression construct, the *rrp41* coding region plus 0.5-kb and 0.65-kb of 5' and 3' genomic sequences, respectively, was PCR-amplified using genomic DNA, and the PCR product was cloned into the *ade6* integration vector pFB366 resulting in pFB855. The *rrp41* expression construct bearing a channel-occluding (co) insertion was created by a 2-step procedure beginning with the cloning of a PCR-amplified genomic fragment containing 0.5-kb of *rrp41* 5' UTR sequences and the first 345-bp of the *rrp41* coding region (fragment #1, amino acids 1-62), followed by the cloning of a second DNA fragment containing nucleotides 346-888 of the *rrp41* coding region (amino acids 63-242) plus 0.65-kb of 3' UTR sequences (fragment #2). The DNA sequence encoding for the channel-occluding insertion was included in the 3' primer used to amplify fragment #1. This procedure inserted an exogenous *Pst*I site between nucleotides 345-346 which was subsequently changed to an *Xho*I site by site-directed mutagenesis to generate the final channel-occluding insertion N-GESEGESEGLE-C (pFB955). This insertion is similar in length and in amino acids composition to a previously described insertion known to physically occlude the central channel of the *S. cerevisiae* RNA exosome (Wasmuth et Lima, 2012). The wild-type and channel-insertion Rrp41 expression constructs were confirmed by sequencing. For single integration into the *ade6* locus, pFB366, pFB855 and pFB955 were linearized and transformed into the conditional P_{nmt1} -rrp41 strain (FBY1347). Positive integrants were confirmed by growth selection on EMM agar plate lacking adenine.

RNA preparation and analyses

Total yeast RNA was extracted using the hot-acid phenol method, resolved on agarose-formaldehyde gels or on polyacrylamide-urea gels in case of RNase H experiment and subsequently transferred onto nylon membranes prior to RNAs crosslinking. Prehybridization and overnight hybridization were carried out in Church buffer at 42°C and

65°C for DNA and RNA probes, respectively. Strand-specific RNA probes used for northern analyses were designed to be complementary to ORF sequences. RNA probes were generated by *in vitro* transcription using the T7 riboprobe system (Ambion) and internally labeled with [α - 32 P]-UTP. DNA probes were generated by 5' end labeling of single-stranded oligonucleotides using the T4 polynucleotide kinase and [γ - 32 P]-ATP. For DNA probes, 2X 15 minute washes were done in 2x SSC, 0.1% SDS followed by 2X 5 minute washes in 0.1x SSC, 0.1% SDS; for RNA probes, 3X 15 minute washes were performed in 1x SSC, 0.1% SDS followed by one 15 minute wash in 0.1x SSC, 0.1% SDS. Membranes were exposed using Phosphor Screens followed by visualization and quantification of signals using a Typhoon Trio instrument and ImageQuant TL, respectively (GE Healthcare). Real-time PCR analyses were done as previously describe(Lemay *et al.*, 2010).

RNase H cleavage assays

Total yeast RNA was treated by RNase H as previously described(Lemay *et al.*, 2010) using a mixture containing DNA oligonucleotides complementary to sequences downstream of the snR3 polyadenylation sites (poly(A)), as illustrated in Supplementary Fig. 1c, in order to detect snR3 read-through transcripts. Annealing of these antisense DNA oligonucleotides to 3'-extended snR3 transcripts (3'-ext) resulted in RNA:DNA hybrids that were cleaved by RNase H, releasing a specific RNA fragment that was detected using a strand-specific riboprobe complementary to this RNase H-cleaved product.

Protein analysis

Total cell extracts were prepared by harvesting mid-log cells in ice-cold lysis buffer (50 mM Tris (pH 7.5), 5 mM MgCl₂, 150 mM NaCl and 0.1% NP-40) containing a cocktail of protease inhibitors (PMSF & PLAAC) prior to glass bead-lysis using a Fastprep instrument. Clarified lysates were normalized for total protein concentration using the Bradford protein assay. For affinity purification experiments, 2 mg of total proteins were subjected to immunoprecipitation using 50 μ l of 50% slurry IgG-sepharose beads that were pre-equilibrated in lysis buffer, and gently mixed for 2 hours. The beads were then washed 3

times with 1 ml of lysis buffer, and the bound proteins were eluted in SDS-PAGE sample buffer. Proteins were separated by SDS-PAGE, transferred to nitrocellulose membranes, and analyzed by immunoblotting using a rabbit polyclonal antibody specific to Dis3 (Kinoshita *et al.*, 1991) (B-Bridge, 63-123; 1:500 (v/v) dilution), a mouse monoclonal antibody specific to α -tubulin (Sigma-Aldrich, T5168; 1:1000 (v/v) dilution), and a rabbit polyclonal antibody against protein A (Sigma-Aldrich, P3775; 1:10,000 (v/v) dilution) in case of TAP-tagged proteins. Validation is provided on the manufacturer's website for the antibodies against α -tubulin and protein A. Membranes were then probed with donkey anti-rabbit conjugated to IRDye 800CW (LICOR, 926-32213; 1:15,000 (v/v) dilution) and a goat anti-mouse conjugated to AlexaFluor 680 (Life technologies, A-21057; 1:15,000 (v/v) dilution). Detection of the proteins was performed using an Odyssey infrared imaging system (LICOR).

Chromatin immunoprecipitation (ChIP) assays

ChIP experiments were performed essentially as described (Larochelle *et al.*, 2012). For ChIP-seq experiments, chromatin immunoprecipitation was done as described above with the following modifications. We scaled up the reactions to obtain larger amount of chromatin for library preparation. Chromatin was immunoprecipitated with Pan mouse IgG Dynabeads (Life technologies, 11041) coated with antibody against Rpb1 (8WG16, Convance MMS-126R) or anti-HA 12CA5 (Roche, 11 583 816 001) for the Rpb3-HA strain (Coudreuse *et al.*, 2010). The 8WG16 antibody is known to recognize the C-terminal domain (CTD) of Rpb1 and measures total RNAPII along genes in yeast (Bataille *et al.*, 2012). Control ChIP assays using untagged strains were performed for all experiments.

Transcription Run-on assays

Transcription run-on experiments were performed essentially as described (Gullerova *et al.*, 2008) with some modifications. Briefly, 150 ml of yeast cells cultures were grown in EMM supplemented with appropriate amino acids and 60 μ M of thiamine for 15 hours to an OD₆₀₀ of 0.20-0.25. Cells were collected by centrifugation, washed with ice-cold

water and resuspended in 950 μ l of ice-cold water. *N*-lauryl sarcosine sulfate (sarkosyl) was added to a final concentration of 0.5% and incubated on ice for 20 min. Cells were recovered by low-speed centrifugation and the supernatant was removed. Cells were then resuspended in 60 μ l of 2.5X transcription buffer (50 mM Tris-HCl pH 7.7, 500 mM KCl, 80 mM MgCl₂) containing 2 mM DTT, 2 mM each ATP, CTP and GTP and 100 μ Ci of [α -³²P]-UTP in a final volume of 150 μ l. Transcription was allowed to proceed for 10 min at 30°C before cells were washed with water followed by total RNA extraction. After partial RNA hydrolysis using 0.2M NaOH on ice for 5 min, RNA was neutralized with 0.14M Tris, 0.14M HCl. Single-stranded RNA probes were generated by *in vitro* transcription using a T7 RNA polymerase and 1 μ g of each probe were immobilized on Hybond-XL (GE Healthcare) membranes by slot-blotting. Membranes were pre-hybridized overnight in 50% formamide, 5X SSPE, 5X Denhart's, 0.4% SDS, and 100 μ g/ml salmon sperm DNA at 45°C. Hybridizations were performed for 24 hours at 45°C after the addition of denatured radiolabeled RNA using 12 mm x 75 mm tubes containing 2 ml of 9% dextran sulfate, 50% formamide, 5X SSPE, 1X Denhart's, 0.4% SDS and 100 μ g/ml salmon sperm DNA. Membranes were washed 2-3 times for 5 min in a large volume of 2X SSC, 0.1% SDS at 45°C. Membranes were exposed using Phosphor Screens followed by visualization and quantification of signals using a Typhoon Trio instrument and ImageQuant TL, respectively. Signal values for each probe were corrected for background signal in a *ura4* ORF region, as all strains used in the run-on assays had a deletion of the *ura4* coding region (*ura4-D18* allele).

Growth assays

For growth assays, 50 μ l of appropriated early-log cell suspensions (optical density_{600nm}: 0.02) with episomally maintained pFB789, pFB811 or pFB818 were added to 50 μ l of EMM Leu- media or 50 μ l of EMM Leu- media supplemented with 7.5 μ M (final concentration) of the non-metabolic inducer anhydrotetracycline hydrochloride (AhTET). Samples were prepared in triplicate in Costar 96-well non-coated polystyrene microplate and cell growth was monitored using a PowerWave microplate spectrophotometer reader (BioTek Instruments) using software settings previously described (Toussaint et Conconi,

2006). For spot assays, exponential cell cultures were adjusted to an optical density OD_{600} :1.0, and then 3 μ l of 10-fold serial dilutions of the suspension were spotted on the appropriated plates. Plates were incubated for 3 to 5 days at 30 °C.

Library preparation and Illumina sequencing

Libraries were prepared from total *S. pombe* RNA and sequenced using the Illumina HiSeq technology. Libraries for RNA-seq analysis were generated as described in (Lemieux *et al.*, 2011). For ChIP-seq analyses, libraries were prepared using the NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina kit (New England BioLabs) according to the manufacturer instructions.

RNA-seq data analysis: Raw FASTQ files quality was assessed using the FastQC software (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Raw reads were mapped, processed and raw read numbers per feature were extracted as described (Lemieux *et al.*, 2011). Fission yeast genome annotation as available on the 9th of May 2011 were used throughout this study. Sequencing scores for each base (number of reads mapping to a given nucleotide position divided by total number of mapped reads) were calculated. RNA-seq scores were displayed along genome annotation using the Bioconductor 'tilingarray' package (Huber *et al.*, 2006). Poly(A) cleavage sites were identified for 792 genes using the approach described by Schlackow *et al.* (Schlackow *et al.*, 2013). Briefly, reads containing stretches of at least 4 adenines at their 3' end were selected. Reads containing poly(A) stretches encoded in the genome were filtered out. Poly(A) cleavage-sites less than 6-nt apart were collapsed into one site. Poly(A) sites were assigned to the nearest upstream annotated gene. For each gene, the poly(A) site with the most hits was defined as "main", while the one the further away from the ORF defines the "longest UTR". Signal accumulation around p(A) cleavage sites were calculated as the number of normalized sequence hits per nucleotide over a fixed distance on either side of the main poly(A) site of these 792 genes. Only genes (n=792) for which poly(A) cleavage sites could be mapped in the wt.2_rna sample have been used for this analysis. Data were median-centered for representation. Differential gene expression analyses in coding (ORF for mRNAs and mature transcript for snoRNAs and

snRNAs) and downstream regions (500-nt downstream of TTS as annotated in Pombase (Wood *et al.*, 2012)) were performed using raw mapped reads numbers and the DESeq Bioconductor package (Anders et Huber, 2010) with samples wt.1_rna and wt.2_rna as untreated samples.

ChIP-seq data analysis: Raw reads were mapped and processed as for RNA-seq experiments except that the data were not treated as strand-specific. Each pull-downs (IP) data were normalized with data from input chromatin (IN), as follows. Sequencing scores for each base (number of reads mapping to a given nucleotide position divided by total number of mapped reads) were calculated. The mean sequencing score in a ten nucleotides window centered around each IP nucleotide was computed and divided by the corresponding mean score calculated for the IN sample. Normalized data were then median-centered and used for subsequent analysis. Chip-seq data were displayed along genome annotation using the jBrowse software (Skinner *et al.*, 2009). Read accumulation 500-nt upstream and 500-nt downstream of the p(A) cleavage sites mapped for the aforementioned 792 genes was calculated as for the RNA-seq data. Average gene analyses were performed on normalized data using custom R scripts (flanking region=500-nt split into 10 bins, genes split into 20 bins) and using gene coordinates (including 5' and 3' UTRs when available) as annotated in Pombase (Wood *et al.*, 2012).

Accession code: RNA-seq and ChIP-seq data have been deposited in the ArrayExpress database under accession number E-MTAB-2237.

SUPPLEMENTARY REFERENCES

(associated with supplementary experimental procedures, figures and tables)

- 46 Lemieux, C. et al. A Pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. *Mol Cell* 44, 108-119 (2011).

- 47 Bahler, J. et al. Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* 14, 943-951 (1998).
- 48 Lemay, J. F. et al. The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. *Mol Cell* 37, 34-45 (2010).
- 49 Erler, A., Maresca, M., Fu, J. & Stewart, A. F. Recombineering reagents for improved inducible expression and selection marker re-use in *Schizosaccharomyces pombe*. *Yeast* 23, 813-823, (2006).
- 50 Maundrell, K. Thiamine-repressible expression vectors pREP and pRIP for fission yeast. *Gene* 123, 127-130 (1993).
- 51 Kinoshita, N., Goebel, M. & Yanagida, M. The fission yeast *dis3+* gene encodes a 110-kDa essential protein implicated in mitotic control. *Mol Cell Biol* 11, 5839-5847 (1991).
- 52 Laroche, M., Lemay, J. F. & Bachand, F. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic Acids Res* 40, 10240-10253 (2012).
- 53 Bataille, A. R. et al. A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* 45, 158-170 (2012).
- 54 Gullerova, M. & Proudfoot, N. J. Cohesin complex promotes transcriptional termination between convergent genes in *S. pombe*. *Cell* 132, 983-995 (2008).
- 55 Toussaint, M. & Conconi, A. High-throughput and sensitive assay to measure yeast cell growth: a bench protocol for testing genotoxic agents. *Nature protocols* 1, 1922-1928 (2006).
- 56 Huber, W., Toedling, J. & Steinmetz, L. M. Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* 22, 1963-1970 (2006).
- 57 Schlackow, M. et al. Genome-wide analysis of poly(A) site selection in *Schizosaccharomyces pombe*. *RNA* (2013).
- 58 Wood, V. et al. PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res* 40, D695-699 (2012).
- 59 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* 11, R106 (2010).

- 60 Skinner, M. E., Uzilov, A. V., Stein, L. D., Mungall, C. J. & Holmes, I. H. JBrowse: a next-generation genome browser. *Genome Res* 19 (2009).
- 61 Mata, J. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol* 10, 1407-1414 (2013).
- 62 Schlackow, M. et al. Genome-wide analysis of poly(A) site selection in *Schizosaccharomyces pombe*. *RNA* 19, 1617-1631 (2013).
- 63 Lemay, J. F. et al. The nuclear poly(A)-binding protein interacts with the exosome to promote synthesis of noncoding small nucleolar RNAs. *Mol Cell* 37, 34-45 (2010).
- 64 Lebreton, A., Tomecki, R., Dziembowski, A. & Seraphin, B. Endonucleolytic RNA cleavage by a eukaryotic exosome. *Nature* 456, 993-996 (2008).
- 65 Schaeffer, D. et al. The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat Struct Mol Biol* 16, 56-62 (2009).
- 66 Wasmuth, E. V. & Lima, C. D. Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. *Mol Cell* 48, 133-144 (2012).
- 67 Sigurdsson, S., Dirac-Svejstrup, A. B. & Svejstrup, J. Q. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell* 38, 202-210 (2010).
- 68 Birse, C. E., Lee, B. A., Hansen, K. & Proudfoot, N. J. Transcriptional termination signals for RNA polymerase II in fission yeast. *EMBO J* 16, 3633-3643 (1997).
- 69 Perreault, A., Lemieux, C. & Bachand, F. Regulation of the Nuclear Poly(A)-binding Protein by Arginine Methylation in Fission Yeast. *J Biol Chem* 282, 7552-7562 (2007).
- 70 Lemieux, C. & Bachand, F. Cotranscriptional recruitment of the nuclear poly(A)-binding protein Pab2 to nascent transcripts and association with translating mRNPs. *Nucleic Acids Res* 37, 3418-3430 (2009).
- 71 Huang, Y., Bayfield, M. A., Intine, R. V. & Maraia, R. J. Separate RNA-binding surfaces on the multifunctional La protein mediate distinguishable activities in tRNA maturation. *Nat Struct Mol Biol* 13, 611-618 (2006).

- 72 Larochelle, M., Lemay, J. F. & Bachand, F. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic Acids Res* 40, 10240-10253 (2012).
- 73 Yamanaka, S., Yamashita, A., Harigaya, Y., Iwata, R. & Yamamoto, M. Importance of polyadenylation in the selective elimination of meiotic mRNAs in growing *S. pombe* cells. *EMBO J* 29, 2173-2181 (2010).
- 74 Lemieux, C. et al. A Pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. *Mol Cell* 44, 108-119 (2011).
- 75 Bezanilla, M., Forsburg, S. L. & Pollard, T. D. Identification of a second myosin-II in *Schizosaccharomyces pombe*: Myp2p is conditionally required for cytokinesis. *Mol Biol Cell* 8, 2693-2705 (1997).
- 76 Coudreuse, D. et al. A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in *S. pombe*. *Curr Biol* 20, 1053-1064 (2010).
- 77 Beaudoin, J., Laliberté, J. & Labbé, S. Functional dissection of Ctr4 and Ctr5 amino-terminal regions reveals motifs with redundant roles in copper transport. *Microbiology* 152, 209-222 (2006).

SUPPLEMENTARY TABLES

Supplemental Table S1. *S. pombe* strains used in this study.

Strain	Genotype	Reference
FBY106	h+ ade6-M216 leu1-32 ura4-D18 his3-D1	69
FBY153	h+ ade6-M210 leu1-32 ura4-D18 his3-D1 Pcf11-TAP::kanMX6	70
FBY195	h- ade6-704 ura4- leu1-32 rrp6::kanMX6	71
FBY209	h+ ade6-M210 leu1-32 ura4-D18 his3-D1 Dis3-TAP::kanMX6	63
FBY245	h- ade6-M210 leu1-32 ura4-D18 his3-D1 cid14::ura4	63
FBY621	h+ ade6-M216 leu1-32 ura4-D18 his3-D1 air1::kanMX6	72
FBY636	h90 ade6-M216 leu1-32 dis3-4::kanMX6	73
FBY721	h? ade6? leu1-32 ura4-D18 his3-D1 P81nmt1-dis3::kanMX6	This study
FBY963	h? ade6? leu1-32 ura4-D18 his3-D1 P81nmt1-mtr4::kanMX6	74
FBY966	h+ ade6-M210 leu1-32 ura4-D18 his7-366	75
FBY1347	h? ade6? leu1-32 ura4-D18 his3-D1 P81nmt1-rrp41::kanMX6	This study
FBY1415	h? ade6? leu1-32 ura4-D18 his3-D1 P81nmt1-dis3::kanMX6 pFB366::ade6	This study

FBY1416	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>dis3</i> ::kanMX6 pFB693:: <i>ade6</i>	This study
FBY1424	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>dis3</i> ::kanMX6 pFB702:: <i>ade6</i>	This study
FBY1428	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>dis3</i> ::kanMX6 pFB707:: <i>ade6</i>	This study
FBY1438	h+ <i>ade6</i> -M216 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 Rrp41-TAP::kanMX6	This study
FBY1451	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>dis3</i> ::kanMX6 pFB708:: <i>ade6</i>	This study
FBY1464	h- Rbp3-3HA::kanMX6	76
FBY1472	h- Rbp3-3HA::kanMX6 P81nmt1- <i>dis3</i> ::natMX6	This study
FBY1528	h+ <i>ade6</i> -M210 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his7</i> -366 P81nmt1- <i>dis3</i> ::natMX6	This study
FBY1529	h+ <i>ade6</i> -M210 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his7</i> -366 <i>tfs1</i> ::kanMX6	This study
FBY1530	h+ <i>ade6</i> -M210 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his7</i> -366 <i>tfs1</i> ::kanMX6 P81nmt1- <i>dis3</i> ::natMX6	This study
FBY1550	h+ <i>ade6</i> -M216 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 Csl4-TAP::kanMX6	This study
FBY1555	h+ <i>ade6</i> -M216 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 Rrp4-TAP::kanMX6	This study
FBY1560	h+ <i>ade6</i> -M216 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 pFB366:: <i>ade6</i>	This study
FBY1562	h+ <i>ade6</i> -M216 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 pFB741:: <i>ade6</i>	This study
FBY1563	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 pFB741:: <i>ade6</i> P81nmt1- <i>dis3</i> ::kanMX6	This study
FBY1565	h+ <i>ade6</i> -M216 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 pFB745:: <i>ade6</i>	This study
FBY1566	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 pFB745:: <i>ade6</i> P81nmt1- <i>dis3</i> ::kanMX6	This study
FBY1741	h+ <i>ade6</i> -M216 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 pFB881:: <i>ade6</i>	This study
FBY1743	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 pFB881:: <i>ade6</i> P81nmt1- <i>dis3</i> ::kanMX6	This study
FBY1758	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>rrp41</i> ::kanMX6 pFB366:: <i>ade6</i>	This study
FBY1795	h+ <i>ade6</i> -M210 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 Pcf11-TAP::kanMX6 P81nmt1- <i>dis3</i> ::natMX6	This study
FBY1884	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>rrp41</i> ::kanMX6 pFB855:: <i>ade6</i>	This study
FBY1886	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>rrp41</i> ::kanMX6 pFB955:: <i>ade6</i>	This study
FBY1891	h+ <i>ade6</i> -M216 <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 pFB366:: <i>ade6</i> Csl4-TAP::natMX6	This study
FBY1895	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>rrp41</i> ::kanMX6 pFB855:: <i>ade6</i> Csl4-TAP::natMX6	This study
FBY1897	h? <i>ade6?</i> <i>leu1</i> -32 <i>ura4</i> -D18 <i>his3</i> -D1 P81nmt1- <i>rrp41</i> ::kanMX6 pFB955:: <i>ade6</i> Csl4-TAP::natMX6	This study

Supplemental Table S2. Plasmids used in this study.

Plasmid	Description	Reference
pFB366	Empty <i>ade6</i> chromosomal integration vector	77
pFB693	WT <i>dis3</i> in <i>ade6</i> chromosomal integration vector	This study
pFB702	Endo- (D166N) version of <i>dis3</i> cloned into pFB366	This study
pFB707	Exo- (D516N) version of <i>dis3</i> cloned into pFB366	This study
pFB708	Endo-/Exo- (D166N/D516N) version of <i>dis3</i> cloned into pFB366	This study
pFB741	<i>ura4</i> reporter system cloned into pFB366	This study
pFB745	<i>ura4</i> reporter system (+332 nts of <i>snR99</i> terminator) cloned into pFB366	This study
pFB789	Empty <i>leu1</i> integrative vector with tet-responsive promoter and tet repressor	This study
pFB811	<i>tfs1</i> cloned into pFB789	This study
pFB818	Mutant version of Tfs1 (D274A/E275A) cloned into pFB789	This study
pFB855	WT <i>rrp41</i> in <i>ade6</i> chromosomal integration vector	This study
pFB881	pFB741 with deletions of the <i>ura4</i> DSE and SDE1 elements and mutation of the SDE2 element	This study
pFB955	<i>rrp41</i> bearing a channel-lining insertion cloned in pFB366	This study

DISCUSSION

Les résultats générés au cours de mes études graduées ont servi à la concoction de trois articles. Une discussion de ces résultats fut préalablement effectuée sur la base des connaissances de l'époque où chaque article fut réalisé. Ici, je discuterai à nouveau des résultats, mais cette fois à la lumière des connaissances actuelles. Je soulèverai également de nouvelles perspectives de recherche et proposerai des façons pour les tester.

Article 1

L'influence du type de polyadénylation sur l'expression génique

La présence d'une queue poly(A) à l'extrémité 3' d'un ARNm confère à ce dernier des avantages non négligeables en lien avec son expression. Notamment, la queue poly(A) d'un ARNm influence positivement la capacité de ce dernier à être traduit en plus d'accroître substantiellement sa stabilité cellulaire (Chen et Shyu, 2011, Preiss et Hentze, 1998). Paradoxalement, l'ajout d'une séquence d'adénosines peut également servir à cibler un ARN vers sa dégradation. Une telle fonction fut initialement identifiée chez la bactérie *E. coli* où l'adénylation du petit ARN non-codant RNAI par PAP I promeut sa dégradation via l'activité 3'→5' exoribonucléase de la PNPase (Xu et Cohen, 1995, Xu *et al.*, 1993) et fut subséquemment démontrée pour de nombreux autres types d'ARNs, tant codants que non-codants, faisant ainsi de la polyadénylation une facette importante du métabolisme d'ARN chez cette bactérie (Li *et al.*, 1998, Mohanty et Kushner, 2006, O'Hara *et al.*, 1995). Analogiquement, chez *S. cerevisiae*, la polyadénylation médiée par le complexe TRAMP, lequel génère une courte queue poly(A) en comparaison de celle ajoutée par la poly(A) polymérase canonique, est intimement liée à la dégradation d'une variété de transcrits par l'exosome à ARN (Schmidt et Butler, 2013). Le concept couramment accepté veut donc que la taille de la queue poly(A) influence le recrutement de protéines stabilisatrices telles que les PABPs. Au noyau, PABPN1/Pab2 s'associe aux queues poly(A) des ARNms. Cette liaison nécessite la présence de 11-14 adénosines, une taille nettement supérieure à celle dérivée de l'activité du complexe TRAMP qui est de 4-5 adénosines (Jia *et al.*, 2011, Meyer

et al., 2002, Wlotzka *et al.*, 2011). Ainsi l'union entre une PABP et une queue poly(A) serait associée à des effets bénéfiques alors qu'un événement de polyadénylation privant l'accès à une PABP serait lié à des effets délétères. La mise en lumière du rôle de Pab2 lors de la biogénèse des *snoRNAs* vient toutefois s'opposer à un tel concept. Rappelons qu'une délétion de *pab2* (*pab2Δ*) entraîne l'accumulation nucléaire de long *snoRNAs* poly(A)⁺ (précurseurs) lesquels sont pris en charge par l'exosome à ARN afin de générer des *snoRNAs* matures poly(A)⁻. Or, cette situation est singulière, car bien que Pab2 soit associée à un effet bénéfique au niveau cellulaire, à savoir la génération de *snoRNAs* matures, sa présence ne sert pas à restreindre l'accès à l'exosome à ARN, mais, au contraire, sert à promouvoir son activité. Cette situation n'est toutefois pas unique à *S. pombe*. Ainsi, chez *S. cerevisiae*, Nab2, une PABP nucléaire, fait partie intégrante d'un mécanisme de surveillance d'ARN lié à l'exosome à ARN (Schmid *et al.*, 2012) alors que chez l'humain PABPN1 et l'exosome à ARN collaborent à la régulation de l'expression d'une variété d'ARNs. (Beaulieu *et al.*, 2012, Bresson et Conrad, 2013, Bresson *et al.*, 2015).

La polyadénylation lors de la biogénèse des *snoRNAs*

Au contraire des ARNm, les *snoRNAs* matures sont dépourvus d'une queue poly(A). À bien des égards, l'ajout d'une telle structure à un *snoRNA* peut paraître futile. Les *snoRNAs* sont strictement localisés au noyau et ne sont donc pas sujet à l'effet positif de la queue poly(A) sur le transit nucléo-cytoplasmique. Les *snoRNAs* sont également des ARNs non-codants et *de facto* l'influence de la queue poly(A) sur le processus de traduction ne peut s'appliquer. Cela dit, les *snoRNAs* sont transitoirement polyadénylés au cours de leur cycle de biogénèse autant chez la levure que chez l'humain (Berndt *et al.*, 2012, Grzechnik et Kufel, 2008). Chez *S. cerevisiae* les *snoRNAs* sont notamment générés à partir de longs précurseurs dont la maturation finale dépend de l'activité de polyadénylation de Trf4 (complexe TRAMP) et de Pap1 (poly(A) polymérase canonique) (Grzechnik et Kufel, 2008). Chez l'humain, les *snoRNAs* sont retrouvés majoritairement à l'intérieur d'introns et dépendent donc d'un événement d'épissage pour leur expression. Suite à l'excision de l'intron, l'extrémité 3' mature du *snoRNA* est générée suite à une cascade d'événements impliquant notamment une étape d'adénylation par PAPD5, l'orthologue présumé de Trf4

(Berndt *et al.*, 2012). La biogénèse des *snoRNAs* chez *S. pombe* implique également la contribution de l'activité de polyadénylation du complexe TRAMP et de la poly(A) polymérase canonique (Pla1). La biogénèse des *snoRNAs* chez *S. pombe* s'est toutefois complexifiée. Alors que chez *S. cerevisiae* le complexe TRAMP et Pap1 collaborent à la maturation des *snoRNAs* (Grzechnik et Kufel, 2008), chez *S. pombe* la polyadénylation médiée par le complexe TRAMP et Pla1 sont associées à une finalité différente au niveau du pré-*snoRNA*. Ainsi, nos données impliquent la participation de Pla1 dans une voie de maturation Pab2/Rrp6-dépendante alors que le complexe TRAMP est lié à une voie de dégradation Dis3-dépendante (article 1) (Larochelle *et al.*, 2012).

Pab2 : recruteur ou activateur de l'exosome à ARN?

Plusieurs de nos données supportent la présence d'une étroite relation entre Pab2 et l'exosome à ARN. Notamment, Pab2 interagit physiquement avec Rrp6/Dis3 *in vivo* et avec Rrp6 *in vitro*. En prenant en considération la capacité de Pab2 à lier une séquence de polyadénosines (Perreault *et al.*, 2007), ces données suggèrent un modèle où la liaison de Pab2 à des substrats polyadénylés mène au recrutement de Rrp6. Cela dit, ce modèle n'a toutefois jamais été testé expérimentalement. En théorie et selon la prémisse que Pab2 agit comme agent de recrutement, l'association de Rrp6 à des substrats polyadénylés, tels que les pré-*snoRNAs*, devrait être modulée en fonction de la présence ou l'absence de Pab2. Une expérience de RIP (*RNA immunoprecipitation*) dirigée contre Rrp6 pourrait être envisagée afin de tester cette hypothèse. Un tel essai nécessiterait toutefois l'utilisation d'une version de Rrp6 dépourvue de son activité 3'→5' exoribonucléase. L'obtention d'un tel mutant chez *S. pombe* ne devrait pas être problématique dans la mesure où les acides aminés nécessaires à cette activité sont conservés (Januszyk *et al.*, 2011). Compte tenu que la mutation de l'activité exoribonucléase n'influence pas la capacité intrinsèque de Rrp6 à lier un substrat d'ARN (Makino *et al.*, 2015), l'utilisation du mutant Rrp6_{exo-} devrait ainsi maximiser la dynamique d'association de Rrp6 aux pré-*snoRNAs* en diminuant la fréquence de dissociation de Rrp6 au substrat laquelle est normalement influencée par la nature distributive de l'activité enzymatique de Rrp6 (Burkard et Butler, 2000). L'utilisation d'un tel mutant est également nécessaire due à la faible présence de pré-*snoRNAs* poly(A)⁺ dans une condition

où Pab2 et Rrp6 sont toutes deux présentes (condition WT). Le mutant Rrp6_{exo-} servirait donc à assurer une quantité non critique de pré-*snoRNAs* poly(A)⁺ maximisant en conséquence le potentiel de recouvrement de ces ARNs. L'isolation d'une quantité moindre de pré-*snoRNAs* poly(A)⁺ dans un contexte Rrp6_{exo-/*pab2*Δ} versus Rrp6_{exo-} supporterait un rôle pour Pab2 dans le recrutement, direct ou indirect, de Rrp6. À l'inverse, l'isolation d'une quantité similaire de pré-*snoRNAs* poly(A)⁺ en présence ou en l'absence de Pab2 impliquerait un rôle alternatif pour Pab2. Chez *S. cerevisiae*, Rrp6 peut être dans un mode dit de liaison ou de dégradation d'ARN (Makino *et al.*, 2015). Or, une interaction entre Pab2 et Rrp6 pourrait induire un changement allostérique chez Rrp6 favorisant l'accès au site catalytique. Des essais de dégradation *in vitro* d'un substrat poly(A)⁺ pourraient notamment être envisagés afin d'éclaircir ce point. De tels essais nécessiteraient toutefois l'utilisation d'un mutant de Pab2 dépourvue de sa capacité à lier l'ARN (Lemieux *et al.*, 2011), mais toujours apte à interagir avec Rrp6. Ainsi, dans l'optique où Pab2 influence l'activité enzymatique de Rrp6 au travers d'un changement structurel, un accroissement de la cinétique de dégradation devrait être observé dans le contexte où Pab2 et Rrp6 sont présents comparativement à une réaction de dégradation impliquant uniquement Rrp6. Si cette conclusion s'avérait fondée, il serait dès lors intéressant d'identifier les domaines chez Pab2 et Rrp6 impliqués dans l'établissement de cette interaction et d'analyser subséquemment par cristallographie les changements structuraux chez Rrp6 induits par la liaison de Pab2.

Cela dit, une double fonction pour Pab2 à l'égard de Rrp6 ne peut pas être écartée. Pab2 pourrait dans un premier temps servir au recrutement de Rrp6 pour ensuite moduler positivement son activité enzymatique. Ce modèle serait en accord avec la fonction connue de Pab2 dans la liaison des queues poly(A) (Perreault *et al.*, 2007) et pourrait expliquer l'accumulation similaire de pré-*snoRNAs* poly(A)⁺ entre des cellules *pab2*Δ et *rrp6*Δ. Bref, les expériences proposées ci-dessus devraient être en mesure d'apporter de nouvelles informations sur le rôle de Pab2 lors du processus de maturation de pré-*snoRNAs* poly(A)⁺.

Une fonction gène-spécifique pour Pab2

En principe, tout ARN pourvu d'une queue poly(A) suffisamment longue peut être associé à Pab2. Considérant que chez *S. pombe* la taille moyenne des queues poly(A) des ARNms est de 30 nucléotides (Subtelny *et al.*, 2014) et que seulement 11-14 adénosines sont requises afin de lier une molécule de PABPN1/Pab2 (Meyer *et al.*, 2002), Pab2 possède ainsi le potentiel d'influencer le métabolisme des ARNs de manière généralisée. Cependant, nos données indiquent qu'un nombre limité d'ARNs ont une expression modulée en absence de Pab2. Outre les *snoRNAs*, ces transcrits incluent des ARN méiotiques, des rétrotransposons, certains ARNms ainsi que des transcrits d'origine cryptiques instables (CUTs) (Zhou *et al.*, 2015). De manière intéressante, la régulation de l'expression de ces ARNs requiert l'activité de la poly(A) polymérase canonique Pla1 (Lemieux *et al.*, 2011, Yamanaka *et al.*, 2010, Zhou *et al.*, 2015). Comment ces transcrits se distinguent-ils alors de tous les autres ARNs qui ont également recours à l'activité de Pla1 lors de leur biogénèse?

Dans le cas des ARNms méiotiques, leur dégradation par l'exosome à ARN en condition de croissance végétative requiert la protéine Mmi1, laquelle reconnaît un motif d'ARN spécifique (DSR : *determinant of selective removal*) (Harigaya *et al.*, 2006), et de nombreuses autres protéines lesquelles comprennent Pab2 (St-Andre *et al.*, 2010, Yamanaka *et al.*, 2010), Red5 (Sugiyama *et al.*, 2013), Iss10 (Sugiyama *et al.*, 2013), Red1 (Sugiyama et Sugioka-Sugiyama, 2011) ainsi que Mtl1 (Egan *et al.*, 2014, Lee *et al.*, 2013, Zhou *et al.*, 2015). Plusieurs de ces facteurs sont également impliqués dans la dégradation de pré-ARNms pourvus de site d'épissage sous-optimaux (Pab2, Mmi1, Mtl1) (Kilchert *et al.*, 2015, Lee *et al.*, 2013, Lemieux *et al.*, 2011). Les *snoRNAs*, pour leur part, sont restreints au noyau suite à leur transcription. Cette rétention nucléaire dépend des séquences nucléotidiques caractéristiques des *snoRNAs* lesquels sont nécessaires à l'assemblage du *snoRNP*. La rétention nucléaire semble donc être une condition nécessaire à la prise en charge d'un transcrit par l'exosome à ARN. La rétention nucléaire expliquerait pourquoi des ARNs arborant une caractéristique similaire, à savoir la présence d'une queue poly(A) Pla1-dépendante, peuvent être liés à des finalités différentes (export nucléaire pour la majorité des ARNms versus rétention nucléaire pour certaines classes d'ARNs). Un point de contrôle

s'exerce donc avant l'acquisition de la compétence à l'export (Tuck et Tollervey, 2013). Selon cette prémisse, maintenir artificiellement un ARNm au noyau devrait induire sa déstabilisation. Il serait ainsi intéressant d'analyser le comportement d'un ARN chimère composé d'un *snoRNA* et partie proximale suivie d'un ARNm en partie distale (promoteur_{*snoRNA*}-*snoRNA*+ARNm-terminateur_{ARNm}). Dans ce contexte, l'ARN chimère serait retenu au noyau par l'entremise des caractéristiques intrinsèques du *snoRNA/snoRNP*, créant ainsi une fenêtre temporelle permettant sa prise en charge par la machinerie de dégradation nucléaire. En théorie, la délétion de *pab2* ou *rrp6* devrait entraîner l'accumulation de transcrits hybride poly(A)⁺ alors qu'en condition WT un ARN correspondant au *snoRNA* ou une version poly(A)⁻ du chimère devrait accumuler conséquence de la dégradation Pab2/Rrp6-dépendante de la partie distale de l'hybride. Or, il pourrait arriver que cet ARN soit insensible à l'action de Pab2/Rrp6. Si une telle situation se présentait, l'explication la plus plausible serait que le 3' UTR de l'ARNm utilisé ne possède pas les déterminants requis pour conférer une susceptibilité à Pab2/Rrp6. Un tel résultat suggèrerait par la même occasion que Pab2 n'est pas le vecteur de recrutement de Rrp6, une conclusion qui se confirmerait suite à l'analyse de l'association de Pab2 aux ARNs chimères poly(A)⁺.

Pab2 et ses partenaires dans la biogénèse des *snoRNAs*

Les partenaires d'interactions de Pab2 furent préalablement identifiés par spectrométrie de masse suite à une purification d'affinité par tandem (Lemieux et Bachand, 2009). Parmi les protéines identifiées, la seule protéine d'intérêt en lien avec la biogénèse des *snoRNAs* correspond à Mtr4, l'hélicase 3'→5' du complexe TRAMP. Cela dit, nos données indiquent que Pab2 et Mtr4 ont des rôles distincts en regard de la biogénèse des *snoRNAs* alors que Pab2 et Mtr4 sont respectivement lié à la voie de maturation et de dégradation. Le réseau d'interactions protéiques de Pab2 a également été caractérisé par double hybride (Vo *et al.*, 2016). Les résultats de cette analyse indiquent que Pab2 interagit avec elle-même ainsi qu'avec Red5, Rmn1, Nab2 et *SPAC328.05* confirmant au passage les résultats de plusieurs études préalables. Ainsi deux molécules de Pab2 peuvent s'associer physiquement *in vivo* tel qu'observé lors d'essais d'immunoprécipitation (Perreault *et al.*,

2007). Pab2 et Nab2, bien qu'elles ne furent point démontrées pour interagir physiquement, participent toutes deux au métabolisme des *snoRNAs* (Grenier St-Sauveur *et al.*, 2013). Red5 fut initialement identifié comme un facteur impliqué dans la dégradation des ARNms méiotiques et co-précipite avec Pab2 (Sugiyama *et al.*, 2013). Rmn1, pour sa part, comprend un domaine de liaison à l'ARN et est impliqué dans l'export des ARNms (Cho *et al.*, 2012). Finalement, *SPAC328.05* est l'orthologue présumé de Gbp2 et Hrb1 de chez *S. cerevisiae*, deux protéines impliqués dans un mécanisme de surveillance des pré-ARNms (Hackmann *et al.*, 2014). Pab2, Red5 et Rmn1 font notamment partis du complexe MTREC (Egan *et al.*, 2014, Lee *et al.*, 2013, Zhou *et al.*, 2015). Ce complexe est un régulateur important d'un nombre important d'ARNs nucléaires chez *S. pombe*. MTREC interagit physiquement avec l'exosome à ARN et avec plusieurs protéines liant l'ARN. Ce complexe se compose d'un noyau central formé des protéines Mtl1, un orthologue présumé de Mtr4, et Red1 et s'associe à 3 sous-modules : Red5-Pab2-Rmn1, Mmi1-Iss10, Ars2-Cbc1-Cbc2. La poly(A) polymérase canonique Pla1 est également associée au complexe MTREC (Egan *et al.*, 2014, Lee *et al.*, 2013, Zhou *et al.*, 2015). De manière intéressante, outre Pab2 et Rrp6, des mutations chez Red1 et Mtl1 entraîne l'accumulation de pré-*snoRNAs* poly(A)⁺ (Lee *et al.*, 2013). Or, il serait pertinent d'analyser si les autres constituants des sous-modules du complexe MTREC participent également à la biogénèse des *snoRNAs*.

L'identification de partenaires protéiques c'est grandement amélioré avec l'émergence de techniques favorisant le maintien des interactions transitoires lors des étapes de purification (Trahan *et al.*, 2016). À cet égard, il serait intéressant de réévaluer le réseau d'interaction protéique de Pab2 afin d'identifier de nouveaux acteurs liés à la biogénèse des *snoRNAs*. Cependant, cette approche pourrait s'avérer exhaustive dans le cas où Pab2 interagirait avec un nombre élevé de protéines. Or, alternativement, une version modifiée d'un *snoRNA* pourrait être créée de manière à insérer en son extrémité 5' des séquences tiges-boucles pouvant être reconnues par la protéine MS2 (Yoon et Gorospe, 2016). L'utilisation d'une telle approche permettrait ainsi une purification exclusive du contenu ribonucléoprotéique associé au *snoRNA*. Cela dit, dans une condition où la biogénèse des *snoRNAs* est perturbée, par exemple dans un contexte où Rrp6 et Dis3 sont toutes deux

absentes (*rrp6Δ/P_{nmt1}-dis3* + thiamine), il est attendu que le contenu protéique associé au *snoRNA* soit enrichi pour des facteurs impliqués dans sa biogénèse.

Les mécanismes entourant la biogénèse des *snoRNAs* chez *S. pombe*

Nos données indiquent que les *snoRNAs* sont sujets à l'action d'une voie de maturation Pab2/Rrp6-dépendante ainsi qu'à une voie de dégradation TRAMP/Dis3-dépendante. Comment ce choix s'effectue-t-il? Une possibilité serait que le choix entre la maturation et la dégradation relève de l'assemblage adéquate du *snoRNP*. Ce concept n'est pas inusuel. En fait chez *S. cerevisiae*, Rrp47, un cofacteur de Rrp6, interagit avec des composantes du *snoRNP* à l'approche de l'extrémité 3' mature du *snoRNA* et cette interaction entraîne la cessation de l'activité 3' → 5' exoribonucléase de Rrp6 (Costello *et al.*, 2011). Dans ce cas, le choix entre la maturation et la dégradation est dicté subséquentement à l'initiation de l'activité exoribonucléase. Chez *S. pombe*, il appert toutefois que ce choix s'effectue au moment de l'évènement de maturation 3' des *snoRNAs*. Ainsi, on note une accumulation synergique de pré-*snoRNAs* poly(A)⁺ de taille similaire lorsqu'une délétion de *pab2* (voie de maturation) est combinée à la délétion ou la déplétion de *cid14* et *dis3* respectivement (voie de dégradation) indiquant que le choix entre la maturation et la dégradation a lieu au même endroit. Le fait que ces deux voies ciblent des populations de transcrits spécifiques indiquent des mécanismes de reconnaissance distincts. Également, compte tenu que des pré-*snoRNAs* TRAMP-dépendant peuvent devenir des substrats de Pab2/Rrp6, cela implique une dynamique d'acquisition de caractéristiques. Un modèle attrayant serait lié à l'exposition de séquences d'ARNs au niveau de la région 3' des *snoRNAs* selon l'assemblage adéquat du *snoRNP*. Cette hypothèse dérive du fait que le complexe MTREC se localise au niveau des gènes de *snoRNA* (Lee *et al.*, 2013) par ChIP et que plusieurs composantes de ce complexe ont la capacité de lier l'ARN. Il serait intéressant de caractériser par CRAC les motifs de liaison des constituants du complexe MTREC afin d'évaluer si de tels sites sont présents au niveau des régions 3' des *snoRNAs*. Alternativement, le complexe MTREC pourrait interagir avec un *snoRNP* une fois ce dernier correctement assemblé. En lien avec cette dernière idée, le complexe MTREC co-précipite avec les constituants protéiques des *snoRNPs* de type C/D et H/ACA (Egan *et al.*, 2014).

Dans ces deux options, un défaut d'assemblage du *snoRNP* favoriserait l'activité du complexe TRAMP lequel est recruté co-transcriptionnellement aux loci de *snoRNAs* (Larochelle *et al.*, 2012). Ce modèle de travail aurait le mérite d'expliquer pourquoi des pré-*snoRNAs* ciblés par le TRAMP/Dis3 peuvent, lorsque la voie de dégradation est modulée négativement, redevenir des substrats Pab2(MTREC)/Rrp6-dépendants. En théorie, si l'assemblage du *snoRNP* est le déterminant sous-jacent à la décision de diriger un *snoRNA* vers la dégradation ou la maturation, seule la voie de dégradation devrait être utilisée dans le cas où un *snoRNA* est créé de manière à empêcher son repliement. Les *snoRNAs* adoptent une structure typique composée de structure tige-boucle. Or, la mutation de cette structure devrait affecter l'association des protéines constituant le *snoRNP*. Ce *snoRNA* mutant ne devrait donc pas être sujet à l'action de Pab2, Rrp6 et de Pla1 et ne devrait pas en conséquence accumulé sous une forme mature.

L'implication des complexes MTREC et TRAMP dans la biogénèse des *snoRNAs* est particulièrement intéressante compte tenu de leur association respective à Rrp6 et Dis3. Chacun de ces complexes possède une hélicase ARN-dépendant 3'→5': Mtl1 pour le complexe MTREC et Mtr4 pour le TRAMP. Chez *S. cerevisiae*, Mtr4 lie une surface d'interaction formée de l'agencement de Rrp47 avec Rrp6 (Schuch *et al.*, 2014). Le domaine par lequel Mtr4 interagit avec cette surface d'interaction est présent chez Mtr4 de *S. pombe*, mais est toutefois absente chez Mtl1 (Zhou *et al.*, 2015). Ainsi, dans le cas d'évènements de dégradation, Mtr4 pourrait diriger les pré-*snoRNAs* poly(A)⁺ vers Rrp6. Cependant, nos données impliquent l'activité de Dis3 dans la voie de dégradation suggérant que les déterminants moléculaires liés au choix de l'activité enzymatique de l'exosome à ARN se sont complexifiés chez *S. pombe* par rapport à *S. cerevisiae*. À cet égard, notons que plusieurs des constituants du complexe MTREC, incluant Pab2, n'ont pas d'équivalent chez *S. cerevisiae*. De plus, il n'est pas exclu que chez *S. pombe* Mtr4 et Rrp6 soient dans l'incapacité d'interagir.

Cela dit, une purification de Mtr4 chez l'humain co-précipite PABPN1 et d'autres orthologues présumés du complexe MTREC (Andersen *et al.*, 2013, Lubas *et al.*, 2011) et PABPN1 est impliqué dans la biogénèse de l'ARN de la télomérase, un ARN aux propriétés

similaires aux *snoRNAs* de type H/ACA (Nguyen *et al.*, 2015). Ces données suggèrent donc que les mécanismes ciblant un ARN vers l'exosome à ARN sont plus similaires entre *S. pombe* et l'humain qu'entre *S. pombe* et *S. cerevisiae*.

Article 2

Divergence fonctionnelle entre Nrd1 et Seb1

Bien qu'ils dépendent de l'ARNPII pour leur synthèse, les *snoRNAs* utilisent chez *S. cerevisiae* un mécanisme de terminaison transcriptionnel qui diffère de celui utilisé au niveau des gènes codants pour des protéines. Ainsi, contrairement aux ARNm, où la terminaison de la transcription est intimement liée au clivage endonucléolytique du transcrit naissant par la machinerie de clivage et de polyadénylation, la terminaison des *snoRNAs*, ainsi que de nombreux autres ARNncs issues de la transcription généralisée du génome, requiert la participation d'un complexe trimérique composé des protéines Nrd1, Nab3 et Sen1 (complexe NNS) (Porrua et Libri, 2015). Cela dit, il est inconnu si un mécanisme de terminaison de type NNS est conservé chez d'autres espèces eucaryotiques. Face à cette problématique, *S. pombe* représente un modèle d'étude de prédilection. Ainsi, tout comme chez *S. cerevisiae*, la majorité des *snoRNAs* sont transcrits à partir d'unité génique indépendante monocistronique (Dieci *et al.*, 2009), la quasi-totalité du génome est transcrit (Dutrow *et al.*, 2008, Rhind *et al.*, 2011, Wilhelm *et al.*, 2008) et des orthologues putatifs pour chacune des protéines du complexe NNS existent (Seb1 \rightarrow Nrd1, Nab3(*S. pombe*) \rightarrow Nab3 et Sen1/Dbp8 \rightarrow Sen1). L'existence présumée d'un complexe NNS chez *S. pombe* repose toutefois essentiellement sur des alignements de séquence et un nombre limité de données biochimiques (Kim *et al.*, 1999, Marina *et al.*, 2013, Mitsuzawa *et al.*, 2003, Moreira *et al.*, 2004, Yu *et al.*, 2013). Curieusement, ces données, bien que peu exhaustives, ont amplement contribué à accepter l'idée qu'un tel mécanisme de terminaison était vraisemblablement présent et fonctionnel chez *S. pombe*.

Arguant en ce sens, nos données lient Seb1 à la terminaison de la transcription des *snoRNAs*. Cependant, et au contraire de la situation prenant place chez *S. cerevisiae* (Steinmetz *et al.*, 2001), la terminaison transcriptionnelle des gènes de *snoRNA* chez

S. pombe ne requiert ni la participation de Nab3, ni celle de Sen1 ou encore celle de Dbl8. Ainsi, alors que l'activité 5' → 3' hélicase de Sen1 est requise pour une terminaison efficace de la transcription des *snoRNAs* et d'ARNs cryptiques chez *S. cerevisiae* (Steinmetz *et al.*, 2006), cette activité enzymatique apparaît *a priori* dispensable chez *S. pombe* (Zhou *et al.*, 2015). On ne peut toutefois exclure un rôle pour Sen1 et/ou Dbl8 dans la terminaison de la transcription d'autres types d'ARN ARNPII-dépendants. À cet égard, Senataxin, une protéine s'apparentant à Sen1 (*S. cerevisiae*) (Moreira *et al.*, 2004), est impliquée, chez l'humain, dans la terminaison de la transcription de gènes codant pour des protéines (Skourti-Stathaki *et al.*, 2011). Cette fonction n'est cependant pas conservée chez *S. cerevisiae* où la déplétion nucléaire de Sen1 entraîne aucun défaut au niveau de la terminaison de la transcription (Schaughency *et al.*, 2014). Cela dit, on ne peut pas écarter la possibilité que Sen1 et/ou Dbl8 soi(en)t fonctionnellement plus similaire(s) à Senataxin que ne l'est Sen1 (*S. cerevisiae*) (Chen *et al.*, 2014). Or, une simple comparaison des profils d'occupation de l'ARNPII à l'échelle du génome entre une souche WT et des souches *sen1Δ*, *dbl8Δ* et *sen1Δ/dbl8Δ* servirait à évaluer le rôle de ces protéines dans la terminaison de la transcription en générale.

L'apparente absence d'un complexe de terminaison de type NNS chez *S. pombe* est également supportée par la présence d'un réseau d'interactions protéiques différent entre Seb1 et Nrd1. Alors que Nrd1 co-purifie majoritairement avec Nab3 et Sen1 (Vasiljeva et Buratowski, 2006), notre analyse de spectrométrie de masse découlant de la purification de Seb1 indique que Nab3, Sen1 et Dbl8 établissent aucune d'interaction stable avec Seb1. À cet égard, le domaine d'interaction avec Nab3 de Nrd1 (acides aminés 151 à 214 inclusivement) ainsi que le domaine d'interaction avec Nrd1 de Nab3 (acides aminés 204 à 248 inclusivement) (Conrad *et al.*, 2000, Vasiljeva *et al.*, 2008) exhibent un faible degré de conservation et de similarité chez Seb1 et Nab3 (*S. pombe*) respectivement. On note que Seb1, Nab3 et Dbl8 sont également absent du réseau d'interaction protéique de Sen1 lequel est essentiellement composé de sous-unités de l'ARNPIII (Legros *et al.*, 2014). Somme toute, ces données sont en accord avec l'absence d'un complexe de terminaison transcriptionnelle de type NNS et pointent vers l'acquisition de fonctions spécialisées pour Seb1, Nab3, Sen1 (Legros *et al.*, 2014, Pan *et al.*, 2012, Sugiyama *et al.*, 2012) et Dbl8 (Yu *et al.*, 2013) chez *S. pombe*. À cet égard, nos données indiquent que Seb1 est fonctionnellement lié au processus

de maturation 3' et de terminaison transcriptionnelle des ARNms. Notamment, Seb1 est enrichi en aval des sites de polyadénylation, co-purifie avec la quasi-totalité des protéines formant la machinerie de clivage et de polyadénylation, est nécessaire au recrutement optimal de ses facteurs en 3' des gènes et sa déplétion entraîne un changement global de la longueur des 3' UTRs, une conséquence de l'utilisation de PAS alternatifs.

Simplification du mécanisme de maturation 3'/terminaison transcriptionnelle chez *S. pombe*

L'absence d'un complexe NNS chez *S. pombe* combiné au lien étroit entre Seb1 et la machinerie de maturation 3' des ARNms et la présence de défauts de maturation 3' et de terminaison transcriptionnelle aux loci de *snoRNAs* en condition de déplétion de Seb1 converge vers l'implication de la machinerie de clivage et de polyadénylation dans la maturation 3' et la terminaison de la transcription des gènes non-codants chez *S. pombe*. En accord avec cette idée, les *snoRNAs* chez *S. pombe* partagent des caractéristiques associées aux ARNms. Ceux-ci sont notamment produits à partir de précurseurs polyadénylés via une voie de maturation nécessitant l'activité de la poly(A) polymérase canonique Pla1 ainsi que la protéine nucléaire liant les queues poly(A) Pab2 (discussion - article 1) (Lemay *et al.*, 2010). De plus, les gènes codants et non-codants possèdent un signal de polyadénylation similaire composé du motif hexamérique canonique humain AAUAAA (Mata, 2013). Ces données indiquent donc, qu'au contraire de *S. cerevisiae*, *S. pombe* utilise un unique mode de maturation 3'/terminaison transcriptionnelle qu'importe la nature de l'ARN transcrit (codant vs non-codant). Il est intéressant de noter qu'une situation similaire est présente chez l'humain où la terminaison de la transcription des *snRNAs* implique les complexes CFI et CFII (O'Reilly *et al.*, 2014). De plus, la terminaison des PROMPTs (*promoter upstream transcripts*), une classe d'ARN non-codant produit en amont du promoteur de gènes actifs, utilise des signaux de polyadénylation similaire à ceux associés aux ARNms (Ntini *et al.*, 2013, Preker *et al.*, 2011). *S. cerevisiae* fait donc figure d'exception dans la manière dont elle promeut la terminaison de la transcription d'ARNs non-codants.

Les déterminants dictant le recrutement de Seb1 au complexe d'élongation

Nos données supportent un modèle où Seb1, par un recrutement initial au niveau de l'ARNPII en élongation par le biais de son domaine CID puis un transfert subséquent au niveau de l'ARN naissant via la reconnaissance de sites de liaison par son domaine RRM, entraîne une maturation 3' et une terminaison transcriptionnelle efficaces. Ces deux domaines présentent des séquences en acide aminé qui sont hautement conservées/similaires à celles retrouvées au niveau des domaines CID et RRM de Nrd1 supportant une certaine conservation de fonction. À cet égard, la terminaison transcriptionnelle médiée par Nrd1 requiert également l'interaction au domaine CTD de l'ARNPII et une liaison à l'ARN naissant. Dans le cas de Nrd1, il est clairement établi que le déterminant majeur de son recrutement au domaine CTD nécessite la présence d'heptades phosphorylés en S⁵ (Vasiljeva *et al.*, 2008). Cette modification post-traductionnelle apparaît tôt au cours du cycle transcriptionnel (Komarnitsky *et al.*, 2000) et par conséquent, le complexe de terminaison transcriptionnel NNS est essentiellement actif au niveau des gènes de petite taille (Gudipati *et al.*, 2008, Heo *et al.*, 2013, Jenks *et al.*, 2008, Steinmetz *et al.*, 2006). Il apparaît toutefois peu plausible que la phosphorylation de la S⁵ soit nécessaire au recrutement de Seb1. Notamment, une version de Seb1 où des résidus conservés, nécessaire chez Nrd1 pour l'interaction au CTD phosphorylé en S⁵ (Ser⁵_{mut}), sont mutés n'engendrent point de changement dans la sélection de PAS ou de défaut au niveau de la terminaison de la transcription. En rétrospective, ces résultats ne sont toutefois pas très étonnants considérant l'enrichissement préférentiel et généralisé de Seb1 à l'extrémité 3' des gènes, une région bien caractérisée pour être associée à la présence d'un CTD phosphorylé en S² (Komarnitsky *et al.*, 2000). Cela dit, outre le chevauchement des profils de localisation à l'extrémité 3' des gènes de Seb1 et de la S² phosphorylée, la liaison de Seb1 au domaine CTD phosphorylé en S² peut également être postulée sur la base des structures de Rtt103, Pcf11 et de SCAF8 (l'orthologue présumé de Seb1 chez les humains) (Becker *et al.*, 2008, Lunde *et al.*, 2010, Meinhart et Cramer, 2004). Rtt103 et SCAF8 lient avec une haute affinité le domaine CTD phosphorylé en S². Cette caractéristique vient de la présence d'une arginine conservée (voir les alignements de séquences dans (Becker *et al.*, 2008)) laquelle établie une interaction de type pont salin avec le groupement phosphate de la S² phosphorylée. Cet acide aminé est

particulièrement important dans la mesure où sa mutation diminue l'affinité de liaison de SCAF8 pour la S² phosphorylé du domaine CTD (Becker *et al.*, 2008) et affecte le recrutement de Rtt103 à l'extrémité 3' des gènes (Lunde *et al.*, 2010). La présence de cette arginine n'est toutefois pas conservée chez Pcf11 ou Nrd1 (l'arginine est remplacée par une asparagine (N) et une méthionine (M) chez Pcf11 et Nrd1 respectivement), deux protéines qui possèdent respectivement peu ou pas d'affinité de liaison pour un domaine CTD phosphorylé en S² (Lunde *et al.*, 2010, Vasiljeva *et al.*, 2008). Chez Seb1, l'acide aminé correspondant est une lysine (K₁₂₄). Tout comme l'arginine, la lysine permet la formation de ponts salins (Errington et Doig, 2005). Il serait donc pertinent d'analyser l'influence d'une mutation de cet acide aminé sur le recrutement de Seb1 à l'extrémité 3' des gènes ainsi que sur les événements de maturation 3' et de terminaison transcriptionnelle. Cette expérience renforcerait le rôle présumé de la phosphorylation de la S² au niveau du recrutement de Seb1.

Cela dit, aucune de nos données ne supporte formellement la présence d'une interaction directe entre Seb1 et le domaine CTD de l'ARNPII malgré que le mutant CID_{mut} de Seb1 le suggère fortement. Le domaine CTD de l'ARNPII adopte une structure β -turn permettant d'accommoder un domaine CID. Cette conformation β -turn est reconnue, indépendamment de l'état de phosphorylation, par des acides aminés du domaine CID via l'établissement d'interactions spécifiques avec le domaine CTD (Becker *et al.*, 2008, Kubicek *et al.*, 2012, Lunde *et al.*, 2010, Meinhart et Cramer, 2004). Les résidus nécessaires pour l'interaction CTD/CID correspondent notamment à ceux mutés au niveau du mutant CID_{mut} de Seb1. Ce mutant abolit notamment le recrutement de Seb1 à l'extrémité 3' des gènes et altère l'utilisation de PAS et la terminaison de la transcription suggérant que Seb1 interagit avec le domaine CTD de l'ARNPII. Par contre, ce mutant ne permet pas de conclure si le recrutement de Seb1 au domaine CTD est phosphorylation-dépendant. Bref, il serait hautement pertinent d'entreprendre des essais de liaisons *in vitro* entre une version recombinante de Seb1 et différentes versions phosphorylées de peptides semblables aux répétitions heptapeptidiques du domaine CTD ainsi que d'élucider la structure de Seb1 par des expériences de cristallographie, afin de mieux comprendre le rôle du domaine CID au niveau du mécanisme de recrutement de Seb1.

D'un autre côté, un mécanisme de recrutement CTD-indépendant ne peut pas être écarté dans le cas de Seb1. En accord avec cette hypothèse, une cristallographie effectuée sur Lsg1, l'homologue de Ctk3 chez *S. cerevisiae*, a révélé une structure ressemblant fortement à celles des domaines CID des protéines Rtt103, Pcf11, Nrd1 et de SCAF8. Cependant, la cavité de liaison au domaine CTD n'est pas conservée et est différemment chargée. En conséquence, le domaine CID prédit de Lgs1 est dans l'incapacité de lier des peptides mimant les répétitions heptapeptidiques du domaine CTD qu'ils soient phosphorylés ou non (Muhlbacher *et al.*, 2015). Ainsi, une possibilité est que Seb1 soit recrutée au niveau du complexe d'élongation à l'extrémité 3' des gènes au travers d'un potentiel domaine CID non canonique effectuant des contacts protéine-protéine autres que l'interaction assumée CID-CTD_{ARNPII}. À cet égard, Nrd1 peut reconnaître une mimique de CTD chez les protéines Trf4 et Mpp6 (Kim *et al.*, 2016, Tudek *et al.*, 2014). Sur cette ligne de pensée, la co-purification de plusieurs sous-unités du complexe de l'ARNPII avec Seb1 est également très intéressante compte tenu que le complexe de l'ARNPII comporte, outre le domaine CTD, plusieurs régions de surface conservées pouvant potentiellement servir à l'établissement d'interactions afin d'optimiser l'activité de l'ARNPII à des stages précis du cycle transcriptionnel incluant la phase de maturation 3'/terminaison de la transcription (Garcia-Lopez et Navarro, 2011). Sur ce point, chez l'humain, l'interaction entre le complexe CPSF et une région autre que le domaine CTD de l'ARNPII est nécessaire à l'induction d'une pause transcriptionnelle et la terminaison de la transcription (Nag *et al.*, 2007). Seb1 interagit notamment avec Rpb7 (Mitsuzawa *et al.*, 2003). Les acides aminés nécessaires à cette interaction ayant été caractérisés, il serait intéressant d'analyser si la perte d'interaction entre Seb1 et Rpb7 entraîne des défauts de maturation 3' et de terminaison transcriptionnelle. Un recrutement de Seb1 par Rpb7 ne pourrait toutefois expliquer le recrutement biaisé de Seb1 à l'extrémité 3' des gènes. Or, notre étude n'établit pas l'ordre par lequel Seb1 est recrutée lors du processus transcriptionnel. Ainsi, il se pourrait que Seb1 soit initialement recrutée par l'ARN pour ensuite interagir avec l'ARNPII via des sous-unités à proximité du canal de sortie de l'ARN (par exemple Rpb7) et/ou le domaine CTD de l'ARNPII (Spahr *et al.*, 2009). En fait, si la liaison à l'ARNPII dépend d'un recrutement initial de Seb1 à l'ARN, l'interaction de Seb1 avec les diverses sous-unités de l'ARNPII devrait disparaître dans le cas d'une purification effectuée avec le mutant RRM (Seb1 RRM_{mut}).

La relation entre Seb1 et la viabilité cellulaire

Seb1 est essentiel pour le maintien de la viabilité cellulaire (Mitsuzawa *et al.*, 2003). Or, les fonctions cellulaires Seb1-dépendantes associées au maintien de la vie sont méconnues. Avant notre identification des fonctions de Seb1 dans les processus de sélection de PAS et de terminaison transcriptionnelle, le seul rôle connu de Seb1 était lié à la formation d'hétérochromatine aux régions péri-centromériques (Marina *et al.*, 2013). Cependant, la perte d'hétérochromatine à ces régions est dispensable pour la viabilité cellulaire. Une possibilité est donc qu'un contrôle serré des 3' UTRs chez *S. pombe* soit requis au maintien de la vie. En fait, l'emplacement où la terminaison de la transcription s'effectue, laquelle est directement conséquent de la longueur du 3' UTR, est d'une importance capitale pour la partition du génome notamment chez les organismes dont le génome est hautement condensé et compartimenté comme celui des levures afin d'éviter, en outre, d'altérer l'expression d'un gène adjacent par le biais d'interférence transcriptionnelle ou d'une collision entre des ARNPs en cours d'élongation (Shearwin *et al.*, 2005). De plus, l'utilisation de PAS (APA) à l'intérieur d'un même 3' UTR (3' UTR-APA) a le potentiel d'influer sur la viabilité cellulaire en affectant la stabilité, l'efficacité traductionnelle et/ou la localisation subcellulaire des isoformes d'ARNms générés (Mayr, 2016).

La régulation de la stabilité d'un transcrite suite à un changement de son 3' UTR est notamment influencée par l'inclusion/exclusion d'éléments en *cis* qui agissent au niveau du recrutement de facteurs séquence-spécifique (Eser *et al.*, 2016, Geisberg *et al.*, 2014, Gupta *et al.*, 2014, Hasan *et al.*, 2014, Mayr et Bartel, 2009, Sandberg *et al.*, 2008). Une comparaison du profil d'expression génique de cellules WT avec celui de cellules ayant subies une déplétion de Seb1 (P_{nmt1}-seb1 + thiamine) indiquent cependant que l'expression des gènes est généralement peu influencée par la déplétion de Seb1. Nos résultats ne peuvent toutefois pas exclure complètement un impact sur la stabilité des ARNs, compte tenu que la stabilité est extrapolée à partir de données statiques qui font abstraction des cinétiques de dégradation et de synthèse. À cet égard, chez *S. cerevisiae*, un système compensatoire existe entre les événements de synthèse et de dégradation (Haimovich *et al.*, 2013, Sun *et al.*, 2013).

On ne peut également pas exclure que la létalité observée suite à la déplétion de *Seb1* soit le résultat d'un changement significatif de l'expression d'une sous-population de gènes essentiels. Néanmoins, le changement limité du profil d'expression des gènes en condition de déplétion de *Seb1* suggère que l'inviabilité de ces cellules n'est pas nécessairement associée à une instabilité accrue des ARNm suite à l'accroissement de leur 3' UTR.

Alternativement, les séquences présentes suite à l'utilisation d'un PAS alternatif peuvent affecter le ciblage approprié d'un ARNm. Chez l'humain, des transcrits dotés d'un long 3' UTR ont notamment tendance à agglomérer au sein de granule à ARN (*RNA granule*) au contraire des mêmes transcrits pourvus d'un 3' UTR court (Han *et al.*, 2012). La présence de telles agrégations cytoplasmiques enrichies en ARN est une caractéristique partagée par de nombreux organismes, incluant *S. pombe* chez qui ces granules d'ARN sont appelées *P-bodies* (Wang *et al.*, 2013). À cet égard, il serait pertinent d'analyser si la déplétion de *Seb1* engendre/accentue la formation de granule à ARN. Les ARNs enrichis au sein de ces granules seraient alors moins disponibles pour être traduits affectant conséquemment la viabilité cellulaire. Dans cette ligne de pensée, chez des lignées cellulaires fibroblastiques embryonnaires de souris, la synthèse protéique de certains ARNm est favorisée par la présence d'un 3' UTR court (Chang *et al.*, 2015). Il serait ainsi pertinent d'analyser si l'association de certains ARNm aux polyribosomes est affectée de manière différentielle entre une condition WT (isoforme court) et une condition où il y a eu une déplétion de *Seb1* (isoforme long).

Finalement, *Seb1* pourrait être impliqué dans l'export nucléaire des ARNm. Cette hypothèse pourrait être facilement testée par une approche d'hybridation à fluorescence *in situ* (FISH) dirigée contre la queue poly(A) des ARNs (sonde oligo d(T)) ou encore sur des gènes candidats en utilisant une sonde ciblant spécifiquement l'isoforme long de l'ARNm. Une telle hypothèse dérive notamment de l'identification chez l'humain d'un rôle pour SRSF7 dans un mécanisme de régulation de l'APA couplé à l'export nucléaire. Comme *Seb1*, SRSF7 est une protéine à domaine RS (région protéique riche en arginine et en sérine) et sa déplétion entraîne l'apparition d'ARNm dotés de 3' UTR plus long. SRSF7 sert de protéine adaptatrice au facteur d'export nucléaire NXF1 (Mitsuzawa *et al.*, 2003, Muller-McNicoll *et*

al., 2016). Toutefois, même si Mex67 (l'orthologue de NXF1 chez *S. pombe*) fut identifié lors de la purification de Seb1, il est difficile d'imaginer comment ce dernier, de par son enrichissement préférentiel en aval des PAS, pourrait directement lier et stabiliser l'association de Mex67 à l'ARNm afin que celui-ci soit exporté. D'un autre côté, les évènements de maturation 3' de l'ARN, de terminaison de la transcription et d'export nucléaire sont fonctionnellement couplés. Ainsi, plusieurs facteurs de maturation 3' tels que Rna14, Rna15 et Pcf11 affectent l'export nucléaire lorsque mutés (Brodsky et Silver, 2000, Hammell *et al.*, 2002). Compte tenu que la déplétion de Seb1 affecte le recrutement de certains facteurs de maturation 3' dont Rna14, on peut soulever l'hypothèse que Seb1 influence l'export nucléaire au travers du recrutement optimal de la machinerie de maturation 3' de l'ARN. À cet égard, chez *S. cerevisiae*, la délétion d'éléments en *cis* nécessaires à la maturation 3' normale du transcrit *cycl* (*cycl-512*) entraîne la génération de transcrits allongés poly(A)⁺ qui accumulent au noyau (Hammell *et al.*, 2002). La ou les raison(s) sous-jacent(s) à cette rétention nucléaire sont toutefois incomprise(s).

Le mécanisme d'action de Seb1 lors de l'évènement de maturation 3' de l'ARN

La polyadénylation alternative peut être modulée par plusieurs mécanismes. L'APA peut notamment être influencée par le niveau protéique de nombreux constituants de la machinerie de clivage et de polyadénylation tels que CstF-64 (Takagaki *et al.*, 1996), CFI-68 (Martin *et al.*, 2012) et PABPN1 (de Klerk *et al.*, 2012, Jenal *et al.*, 2012). L'APA peut également être influencée par l'entremise de nombreuses protéines liant l'ARN (Tian et Manley, 2013), une liste à laquelle peut maintenant se greffer Seb1. De plus, un lien existe entre l'activité transcriptionnelle et le choix d'un PAS (Pinto *et al.*, 2011). Nos données supportent également ce lien dans la mesure où l'utilisation de PAS alternatifs en condition de déplétion de Seb1 est partiellement supprimée lorsque l'élongation de la transcription est défavorisée par l'ajout de 6-AU au milieu de culture. Le concept veut donc que Seb1 module négativement le potentiel d'élongation de l'ARNPII. À cet égard, une possibilité est que la présence de Seb1 est nécessaire au recrutement d'un facteur favorisant la pause de l'ARNPII. Chez l'humain, CPSF-30 fut notamment identifié pour induire un état de pause au niveau du complexe d'élongation (Nag *et al.*, 2007). Or, il est intéressant de souligner que la déplétion

de Seb1 affecte le recrutement de certains facteurs de la machinerie de clivage et de polyadénylation. Il est donc concevable que Seb1 favorise l'utilisation d'un site de polyadénylation proximal via le recrutement de facteurs agissant au niveau de l'ARNPII afin de favoriser une pause au niveau de l'évènement transcriptionnel. À cet égard, il serait intéressant d'analyser le recrutement d'Yth1, l'orthologue présumé de CPSF-30, afin de voir si son recrutement à l'extrémité 3' des gènes est affecté en condition de déplétion de Seb1.

La liaison à l'ARN est une caractéristique importante du mécanisme par lequel Seb1 influence le choix d'un PAS. Il peut être envisagé que la liaison de Seb1 à l'ARN modifie le repliement du pré-ARN naissant favorisant la reconnaissance des éléments en *cis* du PAS par la machinerie de clivage et de polyadénylation. Alors que la vitesse d'élongation de l'ARNPII influence le repliement d'un ARN (Lai *et al.*, 2013), il n'est pas exclu qu'une situation inverse ait lieu où la capacité d'élongation de l'ARNPII est modulée par le repliement de l'ARN en amont. Il se pourrait donc que Seb1 n'induit pas la pause transcriptionnelle à proprement dit, mais plutôt permet de l'accentuer. Or, la structure d'un ARN est directement liée à la durée de la pause, celle-ci étant diminuée ou accentuée selon que l'ARN naissant adopte une structure ou non respectivement. Cette dynamique est notamment liée à l'inhibition du *backtracking* de l'ARNPII en présence d'un ARN structuré (Zamft *et al.*, 2012). Il est intéressant de mentionner que le *backtracking* est prédominant en aval des PAS chez *S. pombe* (discussion - article 3). Ainsi, la liaison de Seb1 à l'ARN naissant pourrait modifier le repliement de l'ARN en aval des PAS de manière à limiter la formation de structures secondaires défavorables au *backtracking* de l'ARNPII. En absence de Seb1, la reprise de l'élongation de la transcription serait rapide, due à la présence de structures secondaires au niveau de l'ARN naissant, créant ainsi une courte période de temps afin que la maturation 3' puisse avoir lieu. Des techniques ont notamment été développées afin de déterminer la structure des ARNms *in vivo* (Ding *et al.*, 2014, Kwok *et al.*, 2015). Il serait ainsi intéressant d'analyser la dynamique du repliement de l'ARN en aval des PAS selon la présence ou non de Seb1.

Nos données indiquent également que plusieurs sites de liaison à Seb1 sont requis afin de promouvoir l'utilisation d'un PAS proximal. Ce résultat fait écho au mécanisme de

terminaison transcriptionnel NNS-dépendant chez qui les motifs de liaison sont fréquemment regroupés afin d'accroître l'efficacité de la terminaison de la transcription (Bacikova *et al.*, 2014, Porrua *et al.*, 2012). Cette efficacité est aussi grandement accentuée par la formation d'un complexe stable entre Nab3 et Nrd1 lequel permet d'augmenter significativement l'affinité aux motifs d'ARN (Carroll *et al.*, 2007, Conrad *et al.*, 2000, Hobor *et al.*, 2011). Nrd1 et Nab3 possèdent tous deux des domaines intrinsèquement désordonnés/de faible complexité (LCD: *low complexity domain*) (O'Rourke *et al.*, 2015). Ces domaines possèdent une capacité intrinsèque à former des agrégations protéiques de type amyloïde pouvant se présenter sous la forme de filament ou encore d'hydrogel (Kato *et al.*, 2012). Les LCDs sont notamment importants lors de la terminaison de la transcription médiée par le complexe NNS (Loya *et al.*, 2013). L'analyse de la séquence protéique de Seb1 par le programme PONDR VL-XT (<http://www.pondr.com/cgi-bin/PONDR/pondr.cgi>) (Romero *et al.*, 2004) prédit de nombreux LCDs. Or, il serait intéressant de déterminer si la fonction de Seb1 dans la sélection de PAS et la terminaison de la transcription implique la polymérisation de molécules de Seb1 par un potentiel domaine LCD. Il serait relativement aisé d'analyser le potentiel d'auto-association de Seb1 par la co-expression en levure de deux versions de Seb1 pourvus d'étiquette de poids moléculaire différent tandis que la formation de structures polymériques pourrait, quant à elle, être analysée par microscopie électronique. Si les expériences susmentionnées s'avéraient positives, une définition du domaine nécessaire à l'auto-assemblage de Seb1 pourrait être entreprise ce qui permettrait ultimement d'analyser le rôle d'une polymérisation de molécules de Seb1 sur la sélection des PAS proximaux et la terminaison de la transcription. Comment cette caractéristique biophysique de Seb1 pourrait-elle toutefois influencer ces processus? Une possibilité est que la combinaison d'une polymérisation de molécules de Seb1 et de la liaison de Seb1 à ces motifs d'ARN «tire» sur l'ARN désalignant l'extrémité 3' du site catalytique de l'ARNPII (phénomène d'*hypertranslocation*) (Peters *et al.*, 2011) favorisant ainsi le *backtracking* de l'ARNPII et donc la pause transcriptionnelle nécessaire à l'accomplissement de l'évènement de maturation 3'. Alternativement, un LCD peut s'associer au domaine CTD de l'ARNPII (Kwon *et al.*, 2013). Ainsi, une deuxième possibilité serait que Seb1 et l'ARNPII s'assemblent sous la forme d'agrégats LCD-dépendants suite à la liaison de Seb1 à ces motifs

d'ARN. Cette structure serait défavorable à l'élongation de la transcription favorisant la maturation 3' de l'ARN et la terminaison transcriptionnelle.

Une conservation de fonction chez les organismes supérieurs?

L'identification d'une fonction pour Seb1 dans le processus d'APA n'est pas insignifiante. La dérégulation de l'utilisation de PAS a notamment été liée à de nombreuses maladies humaines (Curinha *et al.*, 2014). L'orthologue présumé de Seb1 chez l'humain est SCAF8 (Yuryev *et al.*, 1996). Cela dit, la déplétion de SCAF8 n'entraîne pas de défaut de terminaison de la transcription (O'Reilly *et al.*, 2014). Or, les fonctions associées à Seb1 font échos à celles de FUS (*fused in sarcoma*) chez les métazoaires. Tout comme Seb1, FUS possède des domaines CID et RRM et influence l'utilisation de PAS et la terminaison de la transcription par: 1) la liaison à des motifs d'ARN spécifiques, 2) la promotion d'un état de pause au niveau de l'ARNPII et 3) l'interaction avec la sous-unité CPSF-160 du complexe CPSF (Masuda *et al.*, 2015). La manière dont FUS promeut une pause transcriptionnelle est cependant inconnue. Bref, ces données indiquent que les déterminants moléculaires identifiés chez *S. pombe* et liés à la sélection de PAS/terminaison de la transcription ont été conservés chez les organismes supérieurs.

Article 3

L'exosome à ARN nucléaire : un complexe aux multiples fonctions

L'exosome à ARN nucléaire est une machinerie multiprotéique ayant un rôle primordial au sein de la cellule de par son implication dans la dégradation et/ou la maturation de nombreux type d'ARNs. Il est communément accepté que ces activités s'effectuent de manière post-transcriptionnelle. Cela dit, de nombreuses études suggèrent un lien entre la surveillance/maturation d'ARN et l'évènement transcriptionnel. Ainsi, chez *Drosophila melanogaster*, l'exosome à ARN nucléaire est directement associé au complexe transcriptionnel. De plus, des études réalisées chez *S. cerevisiae*, *C. titans*, *D. melanogaster* ou encore les humains ont toutes révélé une association de l'exosome à ARN au niveau de

gènes transcriptionnellement actif (Andrulis *et al.*, 2002, Basu *et al.*, 2011, Hessle *et al.*, 2009, Hieronymus *et al.*, 2004, Lim *et al.*, 2013). Cependant, les détails mécanistiques entourant la contribution de l'exosome à l'ARN dans un processus de surveillance/maturation d'ARN à la chromatine sont rares.

Chez *S. pombe*, nos données indiquent que la déplétion de sous-unités du noyau de l'exosome (Dis3 et Rrp41) entraîne une production généralisée de transcrits dotés d'un 3' UTR allongé (ARN_{UTR+}) lesquels accumulent suite à un défaut de relâchement des complexes d'élongation en aval des PAS. À cet égard, nos données sont en faveur d'un rôle important pour l'exosome à ARN dans un mécanisme lié à la terminaison de la transcription. Normalement, la terminaison de la transcription est étroitement liée au clivage endonucléolytique du pré-ARNm lequel est un prérequis à la terminaison transcriptionnelle de type *torpedo*. Les ARN_{UTR+} qui accumulent suite à la déplétion de Dis3/Rrp41 ne sont toutefois pas sujets à un tel mécanisme de terminaison de la transcription n'ayant pas été clivés au site canonique de maturation 3'. L'exosome à ARN est ainsi engagé dans un mécanisme de terminaison de la transcription «de secours» afin de limiter la progression d'ARNPII en aval des PAS dans les cas d'évènements de clivage non-productifs.

L'importance du *backtracking* de l'ARNPII lors de la terminaison de la transcription médiée par l'exosome à ARN

Parce que la terminaison par l'exosome à ARN requiert l'activité 3' → 5' exonucléase de Dis3, une extrémité d'ARN 3' libre se doit d'être accessible. À cet égard, nos données suggèrent que la génération d'une telle extrémité est liée à la dynamique de l'ARNPII à l'extrémité 3' des gènes. Chez de nombreux organismes, incluant *S. pombe*, l'ARNPII a tendance à s'empiler à l'extrémité 3' des gènes (Anamika *et al.*, 2012, Coudreuse *et al.*, 2010). Cette accumulation se produit suite au passage d'un PAS fonctionnel ce qui entraîne la pause du complexe d'élongation (Nag *et al.*, 2007). De manière importante, nos données révèlent que l'accumulation d'ARNPII à l'extrémité 3' des gènes n'est pas uniquement associée à une pause transcriptionnelle, mais également à un évènement où l'ARNPII recule le long de la matrice d'ADN (*backtracking*). Durant cette marche arrière, le centre catalytique

de l'ARNPII se désengage de l'extrémité terminale de l'ARN naissant causant l'extrusion d'une extrémité 3' d'ARNsb hors du complexe d'élongation (Nudler, 2012). Cette extrémité 3' libre peut alors être prise en charge par l'exosome à ARN. La terminaison exosome à ARN-dépendante peut ainsi être vue comme un modèle de terminaison de la transcription de type *torpedo* inversé compte tenu que l'éviction de l'ARNPII a lieu selon un mécanisme d'action 3' → 5' au lieu de 5' → 3' tel que stipulé par le modèle conventionnel *torpedo*.

Une critique à laquelle s'expose ce modèle est liée au fait que la longueur de l'extrémité 3' libre d'ARN associée à une ARNPII en marche arrière est inférieure aux ~ 31-33 nucléotides nécessaires pour traverser de haut en bas l'exosome à ARN et gagner accès au site catalytique de Dis3 (Bonneau *et al.*, 2009, Makino *et al.*, 2013). En effet, l'identification, au nucléotide près, des extrémités 3' d'ARN associées à l'ARNPII (NET-seq) chez *S. cerevisiae* a révélé qu'en moyenne l'ARNPII recule de 5 à 18 nucléotides (Churchman et Weissman, 2011). Cependant, il est important de mentionner que ces données dérivent d'évènements de pause ayant lieux à l'intérieure des cadres de lecture et en conséquence peuvent diverger de l'étendue du recul de l'ARNPII à l'extrémité 3' des gènes. En fait, la comparaison des données de NET-seq avec celles obtenues par une technique alternative, le PAR-CLIP, indique que le NET-seq a tendance à sous-représenter la présence d'ARNPII à l'extrémité 3' des gènes (Creamer *et al.*, 2011).

Cela dit, des défauts de terminaison transcriptionnelle ont été observés suite à l'obstruction du canal central de l'exosome à ARN indiquant qu'un ARNsb d'une longueur ~ 31-33 nucléotides est nécessairement généré à l'extrémité 3' des gènes. L'explication la plus simple à la présence d'un ARNsb de cette longueur serait que celui-ci dérive de l'évènement de *backtracking*. Alternativement, il est aussi possible que la liaison de l'exosome à ARN à l'extrémité 3' de l'ARNsb favorise des évènements de recul successifs augmentant petit à petit la longueur de l'extrémité 3'. Cependant, cette option n'est pas privilégiée compte tenu de l'interaction relativement faible rapportée entre des ARNsb de moins de 24 nucléotides et l'exosome à ARN (Liu *et al.*, 2014). Or, on ne peut éliminer la possibilité qu'un autre facteur, qui est à identifier, agisse de cette manière afin de fournir à

l'exosome à ARN un substrat d'une longueur adéquate pour être pris en charge par le canal central de l'exosome à ARN.

Les causes potentielles du *backtracking*

De nombreuses études ont révélées que l'élongation de la transcription par l'ARNPII n'est pas un processus uniforme, mais caractérisée par de fréquents évènements de pause et de *backtracking* (Churchman et Weissman, 2011, Gomez-Herreros *et al.*, 2012). Les déterminants qui promeuvent la pause, le recul et la susceptibilité à la terminaison de la transcription exosome à ARN-dépendante sont à être caractérisés, mais pourraient impliquer l'environnement chromatinien. L'environnement des PAS chez *S. pombe* est généralement dépourvu en nucléosome, une caractéristique qui dépend d'une transcription active (Durand-Dubief *et al.*, 2011, Moyle-Heyrman *et al.*, 2013). Toutefois, en aval du PAS la densité en nucléosome s'accroît, une conséquence probable du ralentissement de la cinétique d'élongation de l'ARNPII suite à la transcription du PAS. Cet accroissement de la densité en nucléosome pourrait ainsi servir à l'établissement des évènements de *backtracking* nécessaires à la terminaison médiée par l'exosome à ARN. À l'instar des nucléosomes, d'autres cas de pause transcriptionnelle induite par un blocage au niveau de l'élongation de la transcription ont aussi été rapportés. Chez *S. pombe*, le complexe *cohesin* promeut la terminaison de la transcription à des gènes convergents durant le cycle cellulaire (Gullerova et Proudfoot, 2008), alors que Reb1 agit chez *S. cerevisiae* comme un mécanisme de terminaison de la transcription « de secours » en cas d'une déficience dans l'utilisation d'un PAS en amont (Colin *et al.*, 2014). De manière intéressante, des défauts de terminaison transcriptionnelle sont observables au niveau de gènes convergents en condition de déplétion de l'exosome à ARN et la terminaison de la transcription Reb1-dépendante est liée au *backtracking* de l'ARNPII (Colin *et al.*, 2014). À savoir si l'exosome à ARN est associé aux mécanismes de terminaison de la transcription *cohesin*- et Reb1-dépendants reste toutefois à être déterminé. Les séquences d'ADN sont également une caractéristique importante de la terminaison de la transcription exosome à ARN-dépendante. En fait, lorsque des mutations connues pour promouvoir la pause de l'ARNPII sont introduites à l'extrémité 3' d'un gène modèle (*ura4*), l'exosome à ARN est incapable d'induire la terminaison de la transcription.

Cette donnée suggère que l'environnement génomique à proximité du site de pause peut agir de façon à favoriser l'évènement de *backtracking* de l'ARNPII. En accord avec cette idée, le *backtracking* de l'ARNPII est fortement influencé par la force de l'hybride ARN:ADN à l'intérieur de la bulle transcriptionnelle de l'ARNPII ainsi que par les duplexes ADN:ADN de part et d'autre du complexe d'élongation (Kulish et Struhl, 2001, von Hippel, 1998).

La conservation du mécanisme de terminaison de la transcription de type *torpedo* inversé

L'exosome à ARN est une machinerie protéique conservée au cours de l'évolution et diverses études ont rapporté un lien entre l'exosome à ARN et le processus transcriptionnel chez une variété d'organismes modèles (Andrulis *et al.*, 2002, Basu *et al.*, 2011, Hessle *et al.*, 2009, Hieronymus *et al.*, 2004, Lim *et al.*, 2013, Shah *et al.*, 2014). Chez *S. cerevisiae*, l'exosome à ARN est impliqué dans la terminaison de plusieurs classes d'ARNs non-codants incluant les *snoRNAs* et les CUTs (Gudipati *et al.*, 2012) (Castelnuovo *et al.*, 2013, Fox *et al.*, 2015). Ces types d'ARNs requièrent le complexe NNS pour leur terminaison de la transcription dans un processus intimement lié à leur maturation 3' par l'exosome à ARN. Comme le recrutement des composantes du complexe NNS à un CUT modèle est affecté dans un mutant *RRP6Δ*, il est possible que les défauts de terminaison transcriptionnelle soient la conséquence d'une déficience en complexe NNS. La contribution et le mécanisme par lequel l'exosome à ARN promeut la terminaison de la transcription chez *S. cerevisiae* sont donc plus ou moins clairs.

Chez les métazoaires, une proportion significative des gènes accumule de l'ARNPII quelques nucléotides en aval de leur site d'initiation de la transcription dans un processus connu comme de la pause proximale promoteur-dépendante (Adelman et Lis, 2012). Une étude préalable chez *D. melanogaster* a révélé que les ARNPIIs en pause sont relativement stables malgré une quantité substantielle d'ARNPIIs qui voient leur transcription terminer prématurément (Buckley *et al.*, 2014). De manière intéressante, ces ARNPIIs en pause en région proximale des promoteurs ne peuvent reprendre une élongation de la transcription efficace sans la présence de TFIIS (Adelman *et al.*, 2005). Une telle dépendance envers TFIIS

est un indicateur de la présence d'ARNPIIs en marche arrière ainsi que de la présence d'une extrémité d'ARN 3' libre pouvant potentiellement catalyser la terminaison de la transcription par l'exosome à ARN. En support de ce point, diverses sous-unités de l'exosome à ARN sont présentes aux promoteurs actifs chez *D. melanogaster* (Andrulis *et al.*, 2002, Lim *et al.*, 2013). Toutefois, la déplétion de TFIIS ne réduit pas la densité d'ARNPII en région proximale des promoteurs (Adelman *et al.*, 2005, Henriques *et al.*, 2013), suggérant une contribution relativement mineure de l'exosome à ARN dans la terminaison de la transcription à ces régions. Chez les humains, des défauts de terminaison transcriptionnelle furent rapportés pour le gène encodant le petit ARN nucléaire U2 suite à la déplétion d'une sous-unité du noyau de l'exosome à ARN (Andersen *et al.*, 2013). De plus, le *backtracking* de l'ARNPII est le point d'entrée proposé pour l'exosome à ARN humain durant l'évènement de recombinaison aux loci d'immunoglobulines, une région connue pour induire la pause au niveau de l'ARNPII (Basu *et al.*, 2011). Même si elles sont limitées, ces données suggèrent néanmoins l'existence d'un modèle de terminaison de la transcription de type *torpedo* inversé chez les eucaryotes supérieurs.

La présence d'un mécanisme de terminaison de la transcription de type «double torpedo»

Près du tiers des gènes chez *S. pombe* montre des défauts de terminaison transcriptionnelle importants suite à la déplétion de l'exosome à ARN. À savoir si ces gènes représentent tous des cas où la maturation 3' est inefficace est inconnu. Cependant, comme cette voie de terminaison transcriptionnelle agit sur des ARNPIIs ayant pausées et reculées en aval des PAS, l'exosome à ARN a le potentiel de promouvoir la terminaison de la transcription peu importe si un clivage endonucléolytique a lieu au niveau de l'ARN naissant. Lorsque l'évènement de clivage est productif, l'exosome à ARN et Rat1 (Dhp1 chez *S. pombe*) pourraient alors coopérer selon un modèle de terminaison de la transcription de type «double *torpedo*» afin de dégrader les extrémités 3' et 5' respectivement pour engendrer une terminaison transcriptionnelle optimale. En respect de cette hypothèse, nos données indiquent un profil de dissociation prématuré de l'ARNPII lorsque Tfs1 (TFIIS chez les humains ou Dst1 chez *S. cerevisiae*) est absent de la cellule. En stimulant l'activité

d'hydrolyse intrinsèque de l'ARNPII, laquelle est requise pour réaligner l'extrémité 3' de l'ARN dans le centre catalytique de l'ARNPII en marche arrière, Tfs1/TFIIS promeut la réactivation de la transcription (Izban et Luse, 1992, Reines *et al.*, 1992). Des ARNPIIs en marche arrière en attente de réactivation en aval des PAS sont alors plus susceptibles à demeurer inactives pour une plus longue période de temps dans un mutant *tfs1Δ*. En conséquence, celles-ci représentent une cible plus facile pour l'exoribonucléase 5'→3' de Rat1 (dans le cas d'un clivage productif par la machinerie de maturation 3' de l'ARN) et/ou l'activité exoribonucléase 3'→5' de l'exosome à ARN (dans le cas d'évènements productifs et non-productifs de clivage). En accord avec cette idée, nos résultats indiquent que le profil de terminaison de la transcription prématurée détecté chez le mutant *tfs1Δ* est compensé par la déplétion de l'exosome à ARN. Toutefois, des défauts de terminaison transcriptionnelle sont toujours observables chez le double mutant *tfs1Δ/P_{nmt1}-dis3* lorsque comparé au profil de cellules WT. Une possibilité très intéressante serait que ces défauts soient le reflet d'une terminaison de la transcription sous-optimale due à un mécanisme de type « double *torpedo* » inefficace.

Une façon de tester la présence d'un tel mécanisme de terminaison de la transcription impliquerait l'isolation des ARNs associés à l'ARNPII localisées en aval des PAS. Pour y arriver, une première étape consisterait à purifier l'ARNPII. Pour se faire, la technique du CRAC pourrait être utilisée. Celle-ci a notamment été utilisée pour illustrer des défauts de terminaison transcriptionnelle chez un mutant *dhp1-1* (Tucker *et al.*, 2016). Par contre, la seule purification de l'ARNPII ne permettrait pas de déterminer si les ARNs associés à l'ARNPII ont subi un clivage endonucléolytique au PAS. Une caractéristique importante des transcrits clivés aux PAS est la présence d'une extrémité d'ARN 5' non coiffée au niveau du fragment de clivage en aval. Ainsi, une deuxième ronde de purification devrait avoir lieu afin d'isoler spécifiquement les ARNs non coiffés (Jiao *et al.*, 2008). Le matériel isolé pourrait subséquemment être soumis à un protocole conventionnel de *RNA-seq*. Dans l'optique où un mécanisme «double *torpedo*» est opérationnel dans la cellule, une quantité accrue de séquences (*reads*) en aval des PAS devrait être observée dans le cas du double mutant *dhp1/dis3* en comparaison de chaque simple mutant.

L'INTÉGRATION DES PROJETS DE THÈSE SUR LA CONCEPTION ACTUELLE DU PROCESSUS DE TERMINAISON TRANSCRIPTIONNELLE CHEZ LES EUCARYOTES

Chez les eucaryotes, un défaut lors du processus de terminaison de la transcription peut mener à des conséquences hautement problématiques au niveau de l'expression génique via, en outre, la présence d'interférence transcriptionnelle (gènes en tandem) (Shearwin *et al.*, 2005), l'activation des voies d'interférence à ARN, la collision entre ARN polymérases (gènes convergents) (Gullerova et Proudfoot, 2008, Hobson *et al.*, 2012, Prescott et Proudfoot, 2002) ainsi que la présence d'instabilité génomique suite à la rencontre entre un complexe d'élongation transcriptionnel et la machinerie de réplication de l'ADN (Gaillard *et al.*, 2015). Étrangement, il n'en demeure pas moins que le bagage de connaissance entourant les bases mécanistiques et la régulation de cet évènement fondamental de l'expression des gènes est pour le moins limité.

Une caractéristique importante associée à la terminaison de la transcription médiée par l'ARNPII est l'habileté que possède cet enzyme à voir sa capacité d'élongation être modulée négativement suivant le passage d'un PAS, un phénomène défini comme de la pause transcriptionnelle (voir la section 6.1.4 de l'introduction). Parmi les différents mécanismes promouvant la pause de l'ARNPII, la chromatine via son unité de base, le nucléosome, est probablement le plus commun. La pause de l'ARNPII en aval d'un PAS peut également être induite via l'hybridation du transcrit d'ARN naissant avec le brin d'ADN antisense en amont du complexe d'élongation créant ainsi des structures ARN:ADN dénommé *R-loops* (Skourti-Stathaki et Proudfoot, 2014). De plus, la présence de ces structures est également liée à l'existence de transcription antisense engendrant la formation transitoire d'ARN double brins (ARNdb) et ultimement la création d'îlots localisés de chromatine défavorables au passage de l'ARNPII via l'ajout d'un groupement diméthyl au niveau de la lysine 9 de l'histone 3 (H3K9me2) (Skourti-Stathaki *et al.*, 2014). De plus, la pause transcriptionnelle peut également avoir lieu au travers du recrutement successif et/ou combiné d'un nombre imposant de facteurs au domaine CTD de l'ARNPII puis à l'ARN

naissant tels les membres des complexes de clivage et de polyadénylation. À cette liste peut maintenant se joindre Seb1 (article 2). Indépendamment de son mécanisme d'action, nos données suggèrent un rôle pour Seb1 dans l'accroissement du temps de résidence de l'ARNPII au PAS. Ce faisant, Seb1 sert ainsi à favoriser l'évènement de maturation en 3' de l'ARN et la terminaison de la transcription. Incontestablement, il serait fort intéressant d'évaluer dans un futur rapproché si la pause Seb1-dépendante agit séparément d'autres mécanismes ou si Seb1 fait partie d'un réseau complexe agissant de manière à augmenter et perpétuer la pause de l'ARNPII en aval des PAS.

La terminaison de la transcription en 3' des gènes est un processus hautement régulé et plusieurs ARNms montrent un usage sélectif de différents PAS, un processus décrit comme la polyadénylation alternative (APA) (Tian et Manley, 2013). Plusieurs conditions spécifiques peuvent mener au processus d'APA telles la cinétique transcriptionnelle ou encore la disponibilité en facteurs de clivage et de polyadénylation (Li *et al.*, 2015, Pinto *et al.*, 2011). De plus un nombre grandissant de protéines dotées d'une capacité à lier l'ARN sont connues pour être finement régulées autant sur le plan de l'abondance qu'au niveau de leur localisation cellulaire (Zheng et Tian, 2014). À cet égard, il serait d'intérêt d'évaluer le potentiel de régulation de Seb1 afin de déterminer si la fonction de Seb1 sur le processus d'APA est dynamique.

La maturation PAS-dépendante d'un ARNm est intimement liée à l'éviction de l'ARNPII de la matrice d'ADN. Deux modèles sont couramment utilisés pour expliquer comment la reconnaissance d'un PAS peut promouvoir la terminaison de la transcription à savoir le modèle allostérique et le modèle *torpedo* (Libri, 2015). Chez ce dernier, le clivage de l'ARNm naissant au PAS entraîne l'exposition d'un fragment d'ARN en aval du PAS lequel est physiquement attaché à l'ARNPII. Celui-ci devient dès lors la cible d'une exonucléase 5'→3' qui vient ainsi dégrader progressivement ledit fragment d'ARN et ultimement déloger l'ARNPII de la matrice d'ADN selon un processus de compétition cinétique avec l'ARNPII en cours d'élongation (Fong *et al.*, 2015). Or, mes travaux de thèse ont également mis en à l'avant-scène un nouveau mécanisme de terminaison de la transcription impliquant l'exosome à ARN et délogeant l'ARNPII de la matrice d'ADN selon

une orientation 3'→5' au contraire de l'orientation conventionnelle 5'→3' (*torpedo* inversé) (article 3). Il apparaît donc clair que de nombreux mécanismes, agissant indépendamment ou en conjonction, sont disponibles et utilisés par la cellule afin de contrôler efficacement l'étendue de la transcription par l'ARNPII en région 3' des gènes et ainsi éviter des conséquences catastrophiques sur l'expression génique.

CONCLUSION

Les évènements de maturation 3' de l'ARN et la terminaison de la transcription de l'ARNPII sont deux processus étroitement liés et extrêmement importants pour le fonctionnement adéquat de la cellule. Cela dit, il n'en demeure pas moins que les bases mécanistiques sous-jacents à ces évènements sont encore à ce jour peu comprises et l'objet d'intenses débats idéologiques. Au travers de mes travaux de thèse j'ai ainsi cherché à mieux comprendre ces dits processus via l'utilisation de *S. pombe* comme modèle d'étude (Figure 12), un organisme peut utiliser dans ce domaine de recherche, mais qui a l'avantage de présenter de nombreuses similarités biologiques avec les organismes supérieurs tels les mammifères.

C'est ainsi qu'une fonction fut associée à la protéine Seb1 dans la sélection/promotion de l'évènement de clivage et de polyadénylation au site de polyadénylation (poly(A)) via la génération d'un effet négatif sur la cinétique d'élongation de l'ARNPII (illustré à la figure 12 par une ARNPII à damier) en aval des sites poly(A), un effet médié par le biais d'une série d'interactions avec le domaine CTD de l'ARNPII, l'ARN naissant ainsi que la machinerie de clivage et de polyadénylation (CPF/CFIA/CFIB/Pap1) (article 2). Ce mécanisme d'action fut notamment montré pour être effectif autant au niveau des gènes codants (ARNm) que non-codants (ARNnc (*snoRNA*)) expliquant au passage l'implication de la *poly(A)-binding* nucléaire Pab2 et la machinerie de clivage et de polyadénylation dans la maturation exosome à ARN-dépendante des *snoRNAs* (article 1). Finalement, il fut montré que l'exosome à ARN n'est pas uniquement dédié à des fonctions post-transcriptionnelles, mais que celui-ci peut également agir au niveau co-transcriptionnelle afin de maximiser et/ou de s'assurer d'une éviction efficace de l'ARNPII de la matrice d'ADN selon un mécanisme

impliquant la dégradation d'un fragment d'ARN lié à des ARNPII en état de pause en aval des sites poly(A) (article 3). Combinées, ces découvertes illustrent bien la nature complexe des processus de maturation 3' de l'ARN et de la terminaison transcriptionnelle de l'ARNPII ainsi que le haut degré d'interconnexion entre ces deux processus. Malgré le fait que ces études aient été effectuées chez *S. pombe*, il n'en demeure pas moins que l'ensemble des facteurs protéiques susmentionnés possèdent des orthologues putatifs chez les mammifères suggérant une possible conservation des mécanismes décrits dans ces travaux de thèse. Somme toute, ces travaux ont mis en lumière des facteurs et mécanismes impliqués dans la maturation 3', la terminaison de la transcription et le métabolisme nucléaire d'ARNs ARNPII-dépendants offrant au passage de nouvelles perspectives de recherche pour ces domaines d'étude.

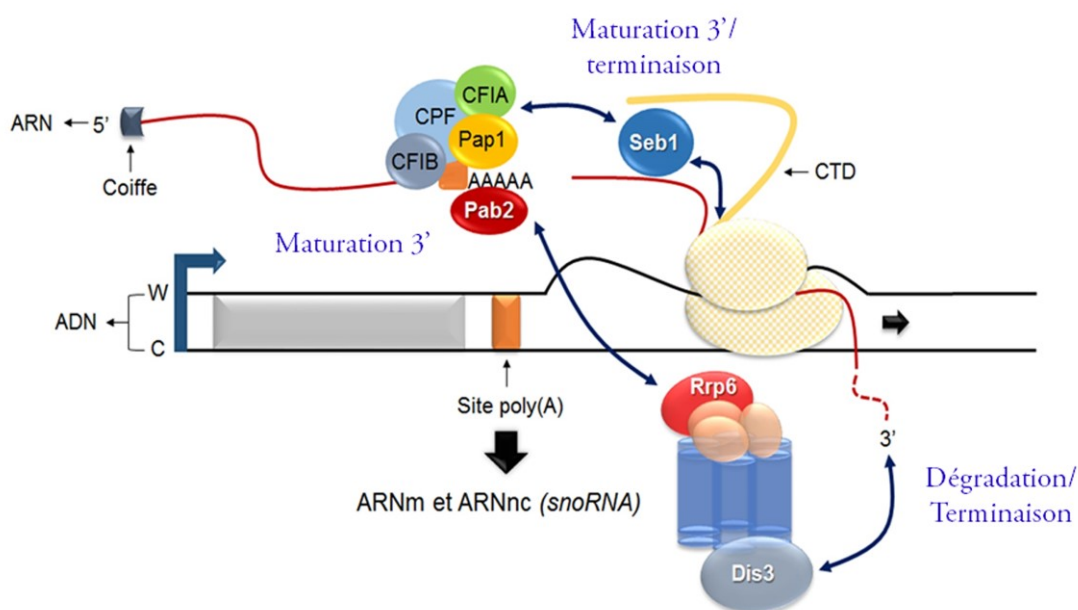


Figure 12. Résumé général des travaux de thèse.

Schéma récapitulatif des avancées scientifiques dérivées des travaux de thèse associées aux événements de maturation 3' de l'ARN et de la terminaison de la transcription chez l'organisme modèle *S. pombe*. Voir la section conclusion pour plus de détails. Le schéma n'est pas à l'échelle.

REMERCIEMENTS

J'aimerais en tout premier lieu remercier François Bachand, mon superviseur durant mes études graduées de maîtrise et de doctorat. François m'a donné l'opportunité de travailler sur des projets stimulants. J'ai toujours senti une énorme confiance et un support constant de la part de François. Merci de m'avoir montré à faire de la recherche inspirée et de m'avoir donné l'opportunité de développer un esprit scientifique qui m'est propre.

Je tiens également à remercier mes amis et collègues présents et passés du laboratoire ainsi que du département de Biochimie avec qui j'ai eu l'opportunité de travailler dans un environnement très agréable et toujours stimulant scientifiquement.

Un merci aux professeurs Simon Labbé, Sherif Abou Elela, et Jacques Côté d'avoir bien voulu être membre de mon comité d'évaluation de thèse. Je tiens aussi à souligner la contribution du CRSNG pour leur important soutien financier lors de mes études graduées.

Mes études graduées n'auront pas été aussi fructueuse sans la précieuse collaboration de mes nombreux co-auteurs. Un énorme merci au laboratoire du professeur Jürg Bähler, du professeur Sander Granneman et du professeur Bin Tian pour la confiance démontrée envers mes projets de recherche. Un merci tout particulier au professeur Samuel Marguerat, Sophie Atkinson, Daniel Lackner, Rob van Nues et Xiaochuan Liu.

J'aimerais de plus remercier ma famille, en particulier mes parents Nicole et Lucien et ma sœur Marie-France, ainsi que ma belle-famille pour le soutien continué durant ce long cheminement. Enfin, j'aimerais dire le plus gros des mercis à Kim. Tu m'as constamment poussé à donner le meilleur de moi-même et as été pour moi une source de réconfort dans les moments plus ardues. Ton support, tes encouragements, ton amitié, ton amour sont ce qu'il y a de plus important pour moi. Je suis fière de mon parcours, mais je suis encore plus fière de l'avoir fait à tes côtés.

LISTE DES RÉFÉRENCES

- Adelman K, Lis JT (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* 13(10): 720-731.
- Adelman K, Marr MT, Werner J, Saunders A, Ni Z, Andrulis ED, Lis JT (2005) Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol Cell* 17(1): 103-112.
- Ahn SH, Kim M, Buratowski S (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* 13(1): 67-76.
- Akhtar MS, Heidemann M, Tietjen JR, Zhang DW, Chapman RD, Eick D, Ansari AZ (2009) TFIIF kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* 34(3): 387-393.
- Allison LA, Moyle M, Shales M, Ingles CJ (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* 42(2): 599-610.
- Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *Embo J* 18(19): 5399-5410.
- Allmang C, Mitchell P, Petfalski E, Tollervey D (2000) Degradation of ribosomal RNA precursors by the exosome. *Nucleic Acids Res* 28(8): 1684-1691.
- Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D, Mitchell P (1999) The yeast exosome and human PM-Scl are related complexes of 3' to 5' exonucleases. *Genes Dev* 13(16): 2148-2158.
- Anamika K, Gyenis A, Poidevin L, Poch O, Tora L (2012) RNA polymerase II pausing downstream of core histone genes is different from genes producing polyadenylated transcripts. *PLoS One* 7(6): e38769.
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10): R106.
- Andersen PR, Domanski M, Kristiansen MS, Storvall H, Ntini E, Verheggen C, Schein A, Bunkenborg J, Poser I, Hallais M, Sandberg R, Hyman A, LaCava J, Rout MP, Andersen JS, Bertrand E, Jensen TH (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat Struct Mol Biol* 20(12): 1367-1376.
- Andrulis ED, Werner J, Nazarian A, Erdjument-Bromage H, Tempst P, Lis JT (2002) The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature* 420(6917): 837-841.

Aranda A, Proudfoot NJ (1999) Definition of transcriptional pause elements in fission yeast. *Mol Cell Biol* 19(2): 1251-1261.

Aravind L, Watanabe H, Lipman DJ, Koonin EV (2000) Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc Natl Acad Sci U S A* 97(21): 11319-11324.

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

Armache KJ, Mitterweger S, Meinhart A, Cramer P (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* 280(8): 7131-7134.

Auffinger P, Westhof E (1997) Rules governing the orientation of the 2'-hydroxyl group in RNA. *J Mol Biol* 274(1): 54-63.

Azuma Y, Yamagishi M, Ueshima R, Ishihama A (1991) Cloning and sequence determination of the *Schizosaccharomyces pombe* rpb1 gene encoding the largest subunit of RNA polymerase II. *Nucleic Acids Res* 19(3): 461-468.

Bachand F, Lackner DH, Bahler J, Silver PA (2006) Autoregulation of ribosome biosynthesis by a translational response in fission yeast. *Mol Cell Biol* 26(5): 1731-1742.

Bacikova V, Pasulka J, Kubicek K, Stefl R (2014) Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res* 42(12): 8024-8038.

Baejen C, Torkler P, Gressel S, Essig K, Soding J, Cramer P (2014) Transcriptome maps of mRNP biogenesis factors define pre-mRNA recognition. *Mol Cell* 55(5): 745-757.

Bahler J, Wu JQ, Longtine MS, Shah NG, McKenzie A, 3rd, Steever AB, Wach A, Philippsen P, Pringle JR (1998) Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* 14(10): 943-951.

Ballarino M, Morlando M, Pagano F, Fatica A, Bozzoni I (2005) The cotranscriptional assembly of snoRNPs controls the biosynthesis of H/ACA snoRNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* 25(13): 5396-5403.

Barabino SM, Ohnacker M, Keller W (2000) Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. *EMBO J* 19(14): 3778-3787.

Barilla D, Lee BA, Proudfoot NJ (2001) Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 98(2): 445-450.

Bartolomei MS, Halden NF, Cullen CR, Corden JL (1988) Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Mol Cell Biol* 8(1): 330-339.

Basu U, Meng FL, Keim C, Grinstein V, Pefanis E, Eccleston J, Zhang T, Myers D, Wasserman CR, Wesemann DR, Januszyk K, Gregory RI, Deng H, Lima CD, Alt FW (2011) The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 144(3): 353-363.

Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A, Bergeron M, Hanes SD, Robert F (2012) A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. *Mol Cell* 45(2): 158-170.

Baxter-Roshek JL, Petrov AN, Dinman JD (2007) Optimization of ribosome structure and function by rRNA base modification. *PLoS One* 2(1): e174.

Beaudoin J, Labbe S (2006) Copper induces cytoplasmic retention of fission yeast transcription factor *cuf1*. *Eukaryot Cell* 5(2): 277-292.

Beaudoing E, Freier S, Wyatt JR, Claverie JM, Gautheret D (2000) Patterns of variant polyadenylation signal usage in human genes. *Genome Res* 10(7): 1001-1010.

Beaulieu YB, Kleinman CL, Landry-Voyer AM, Majewski J, Bachand F (2012) Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. *PLoS Genet* 8(11): e1003078.

Becker R, Loll B, Meinhart A (2008) Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* 283(33): 22659-22669.

Bentley DL (2014) Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* 15(3): 163-175.

Berndt H, Harnisch C, Rammelt C, Stohr N, Zirkel A, Dohm JC, Himmelbauer H, Tavanez JP, Huttelmaier S, Wahle E (2012) Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. *RNA* 18(5): 958-972.

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

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

Berriz GF, Beaver JE, Cenik C, Tasan M, Roth FP (2009) Next generation software for functional trend analysis. *Bioinformatics* 25(22): 3043-3044.

Bienroth S, Keller W, Wahle E (1993) Assembly of a processive messenger RNA polyadenylation complex. *EMBO J* 12(2): 585-594.

Bienroth S, Wahle E, Suter-Crazzolara C, Keller W (1991) Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J Biol Chem* 266(29): 19768-19776.

Birse CE, Lee BA, Hansen K, Proudfoot NJ (1997) Transcriptional termination signals for RNA polymerase II in fission yeast. *EMBO J* 16(12): 3633-3643.

Birse CE, Minvielle-Sebastia L, Lee BA, Keller W, Proudfoot NJ (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science* 280(5361): 298-301.

Bizarro J, Charron C, Boulon S, Westman B, Pradet-Balade B, Vandermoere F, Chagot ME, Hallais M, Ahmad Y, Leonhardt H, Lamond A, Manival X, Branlant C, Charpentier B, Verheggen C, Bertrand E (2014) Proteomic and 3D structure analyses highlight the C/D box snoRNP assembly mechanism and its control. *J Cell Biol* 207(4): 463-480.

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

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

Bowman EA, Kelly WG (2014) RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases. *Nucleus* 5(3): 224-236.

Brais B, Bouchard JP, Xie YG, Rochefort DL, Chretien N, Tome FM, Lafreniere RG, Rommens JM, Uyama E, Nohira O, Blumen S, Korczyn AD, Heutink P, Mathieu J, Duranceau A, Codere F, Fardeau M, Rouleau GA (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat Genet* 18(2): 164-167.

Bresson SM, Conrad NK (2013) The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. *PLoS Genet* 9(10): e1003893.

Bresson SM, Hunter OV, Hunter AC, Conrad NK (2015) Canonical Poly(A) Polymerase Activity Promotes the Decay of a Wide Variety of Mammalian Nuclear RNAs. *PLoS Genet* 11(10): e1005610.

Briani F, Carzaniga T, Deho G (2016) Regulation and functions of bacterial PNPase. *Wiley Interdiscip Rev RNA* 7(2): 241-258.

Brodsky AS, Silver PA (2000) Pre-mRNA processing factors are required for nuclear export. *RNA* 6(12): 1737-1749.

Brown KM, Gilmartin GM (2003) A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* 12(6): 1467-1476.

Buckley MS, Kwak H, Zipfel WR, Lis JT (2014) Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation. *Genes Dev* 28(1): 14-19.

Buratowski S (2003) The CTD code. *Nat Struct Biol* 10(9): 679-680.

Buratowski S (2005) Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol* 17(3): 257-261.

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

Butler JS, Mitchell P (2011) Rrp6, rrp47 and cofactors of the nuclear exosome. *Adv Exp Med Biol* 702(91-104).

Bycroft M, Hubbard TJ, Proctor M, Freund SM, Murzin AG (1997) The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* 88(2): 235-242.

Cahill NM, Friend K, Speckmann W, Li ZH, Terns RM, Terns MP, Steitz JA (2002) Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP. *EMBO J* 21(14): 3816-3828.

Calado A, Kutay U, Kuhn U, Wahle E, Carmo-Fonseca M (2000) Deciphering the cellular pathway for transport of poly(A)-binding protein II. *RNA* 6(2): 245-256.

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

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

Callebaut I, Moshous D, Mornon JP, de Villartay JP (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* 30(16): 3592-3601.

Carlile TM, Rojas-Duran MF, Zinshteyn B, Shin H, Bartoli KM, Gilbert WV (2014) Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515(7525): 143-146.

Carneiro T, Carvalho C, Braga J, Rino J, Milligan L, Tollervey D, Carmo-Fonseca M (2007) Depletion of the yeast nuclear exosome subunit Rrp6 results in accumulation of polyadenylated RNAs in a discrete domain within the nucleolus. *Mol Cell Biol* 27(11): 4157-4165.

Carpy A, Krug K, Graf S, Koch A, Popic S, Hauf S, Macek B (2014) Absolute proteome and phosphoproteome dynamics during the cell cycle of *Schizosaccharomyces pombe* (Fission Yeast). *Mol Cell Proteomics* 13(8): 1925-1936.

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

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

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

Chan SL, Huppertz I, Yao C, Weng L, Moresco JJ, Yates JR, 3rd, Ule J, Manley JL, Shi Y (2014) CPSF30 and Wdr33 directly bind to AAUAAA in mammalian mRNA 3' processing. *Genes Dev* 28(21): 2370-2380.

Chang H, Lim J, Ha M, Kim VN (2014) TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* 53(6): 1044-1052.

Chang JW, Zhang W, Yeh HS, de Jong EP, Jun S, Kim KH, Bae SS, Beckman K, Hwang TH, Kim KS, Kim DH, Griffin TJ, Kuang R, Yong J (2015) mRNA 3'-UTR shortening is a molecular signature of mTORC1 activation. *Nat Commun* 6(7218).

Chao LC, Jamil A, Kim SJ, Huang L, Martinson HG (1999) Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. *Mol Cell Biol* 19(8): 5588-5600.

Chekanova JA, Gregory BD, Reverdatto SV, Chen H, Kumar R, Hooker T, Yazaki J, Li P, Skiba N, Peng Q, Alonso J, Brukhin V, Grossniklaus U, Ecker JR, Belostotsky DA (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the *Arabidopsis* transcriptome. *Cell* 131(7): 1340-1353.

Chen CY, Chang CC, Yen CF, Chiu MT, Chang WH (2009) Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc Natl Acad Sci U S A* 106(1): 127-132.

Chen CY, Shyu AB (2011) Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* 2(2): 167-183.

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

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

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

Chlebowski A, Lubas M, Jensen TH, Dziembowski A (2013) RNA decay machines: the exosome. *Biochim Biophys Acta* 1829(6-7): 552-560.

Cho EJ, Takagi T, Moore CR, Buratowski S (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* 11(24): 3319-3326.

Cho YS, Jang S, Yoon JH (2012) Isolation of a novel *rmn1* gene genetically linked to *spnab2* with respect to mRNA export in fission yeast. *Mol Cells* 34(3): 315-321.

Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469(7330): 368-373.

Colin J, Candelli T, Porrua O, Boulay J, Zhu C, Lacroute F, Steinmetz LM, Libri D (2014) Roadblock termination by *reb1p* restricts cryptic and readthrough transcription. *Mol Cell* 56(5): 667-680.

Connelly S, Manley JL (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* 2(4): 440-452.

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

Corden JL, Cadena DL, Ahearn JM, Jr., Dahmus ME (1985) A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A* 82(23): 7934-7938.

Costello JL, Stead JA, Feigenbutz M, Jones RM, Mitchell P (2011) The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3'-maturation. *J Biol Chem* 286(6): 4535-4543.

Coudreuse D, van Bakel H, Dewez M, Soutourina J, Parnell T, Vandenhoute J, Cairns B, Werner M, Hermand D (2010) A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in *S. pombe*. *Curr Biol* 20(12): 1053-1064.

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12): 1367-1372.

Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ, Jawhari A, Jennebach S, Kamenski T, Kettenberger H, Kuhn CD, Lehmann E, Leike K, Sydow JF, Vannini A (2008) Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* 37(337-352).

Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292(5523): 1863-1876.

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

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6): 1188-1190.

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

Darzacq X, Jady BE, Verheggen C, Kiss AM, Bertrand E, Kiss T (2002) Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J* 21(11): 2746-2756.

Darzacq X, Kittur N, Roy S, Shav-Tal Y, Singer RH, Meier UT (2006) Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *J Cell Biol* 173(2): 207-218.

Davidson L, Muniz L, West S (2014) 3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells. *Genes Dev* 28(4): 342-356.

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

Davis DR (1995) Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res* 23(24): 5020-5026.

de Klerk E, Venema A, Anvar SY, Goeman JJ, Hu O, Trollet C, Dickson G, den Dunnen JT, van der Maarel SM, Raz V, t Hoen PA (2012) Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. *Nucleic Acids Res* 40(18): 9089-9101.

de Vries H, Ruegsegger U, Hubner W, Friedlein A, Langen H, Keller W (2000) Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J* 19(21): 5895-5904.

Decatur WA, Fournier MJ (2002) rRNA modifications and ribosome function. *Trends Biochem Sci* 27(7): 344-351.

Decker CJ, Parker R (1993) A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation. *Genes Dev* 7(8): 1632-1643.

Dengl S, Cramer P (2009) Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* 284(32): 21270-21279.

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

Dez C, Noaillic-Depeyre J, Caizergues-Ferrer M, Henry Y (2002) Naf1p, an essential nucleoplasmic factor specifically required for accumulation of box H/ACA small nucleolar RNPs. *Mol Cell Biol* 22(20): 7053-7065.

Dheur S, Voile TA, Voisinet-Hakil F, Minet M, Schmitter JM, Lacroute F, Wyers F, Minvielle-Sebastia L (2003) Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. *Embo J* 22(11): 2831-2840.

Dichtl B, Blank D, Sadowski M, Hubner W, Weiser S, Keller W (2002) Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. *EMBO J* 21(15): 4125-4135.

Dichtl B, Keller W (2001) Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. *EMBO J* 20(12): 3197-3209.

Dieci G, Conti A, Pagano A, Carnevali D (2013) Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. *Biochim Biophys Acta* 1829(3-4): 296-305.

Dieci G, Preti M, Montanini B (2009) Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics* 94(2): 83-88.

Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505(7485): 696-700.

Dominski Z, Yang XC, Purdy M, Wagner EJ, Marzluff WF (2005) A CPSF-73 homologue is required for cell cycle progression but not cell growth and interacts with a protein having features of CPSF-100. *Mol Cell Biol* 25(4): 1489-1500.

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

Drissi R, Dubois ML, Douziech M, Boisvert FM (2015) Quantitative Proteomics Reveals Dynamic Interactions of the Minichromosome Maintenance Complex (MCM) in the Cellular Response to Etoposide Induced DNA Damage. *Mol Cell Proteomics* 14(7): 2002-2013.

Duncan CD, Mata J (2014) The translational landscape of fission-yeast meiosis and sporulation. *Nat Struct Mol Biol* 21(7): 641-647.

Dupuis-Sandoval F, Poirier M, Scott MS (2015) The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdiscip Rev RNA* 6(4): 381-397.

Durand-Dubief M, Svensson JP, Persson J, Ekwall K (2011) Topoisomerases, chromatin and transcription termination. *Transcription* 2(2): 66-70.

Dutrow N, Nix DA, Holt D, Milash B, Dalley B, Westbroek E, Parnell TJ, Cairns BR (2008) Dynamic transcriptome of *Schizosaccharomyces pombe* shown by RNA-DNA hybrid mapping. *Nat Genet* 40(8): 977-986.

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

Egan ED, Braun CR, Gygi SP, Moazed D (2014) Post-transcriptional regulation of meiotic genes by a nuclear RNA silencing complex. *RNA* 20(6): 867-881.

Egloff S, O'Reilly D, Chapman RD, Taylor A, Tanzhaus K, Pitts L, Eick D, Murphy S (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science* 318(5857): 1777-1779.

Egloff S, Szczepaniak SA, Dienstbier M, Taylor A, Knight S, Murphy S (2010) The integrator complex recognizes a new double mark on the RNA polymerase II carboxyl-terminal domain. *J Biol Chem* 285(27): 20564-20569.

Erler A, Maresca M, Fu J, Stewart AF (2006) Recombineering reagents for improved inducible expression and selection marker re-use in *Schizosaccharomyces pombe*. *Yeast* 23(11): 813-823.

Errington N, Doig AJ (2005) A phosphoserine-lysine salt bridge within an alpha-helical peptide, the strongest alpha-helix side-chain interaction measured to date. *Biochemistry* 44(20): 7553-7558.

Eser P, Wachutka L, Maier KC, Demel C, Boroni M, Iyer S, Cramer P, Gagneur J (2016) Determinants of RNA metabolism in the *Schizosaccharomyces pombe* genome. *Mol Syst Biol* 12(2): 857.

Evguenieva-Hackenberg E, Hou L, Glaeser S, Klug G (2014) Structure and function of the archaeal exosome. *Wiley Interdiscip Rev RNA* 5(5): 623-635.

Fan X, Dion P, Laganier J, Brais B, Rouleau GA (2001) Oligomerization of polyalanine expanded PABPN1 facilitates nuclear protein aggregation that is associated with cell death. *Hum Mol Genet* 10(21): 2341-2351.

Fasken MB, Leung SW, Banerjee A, Kodani MO, Chavez R, Bowman EA, Purohit MK, Rubinson ME, Rubinson EH, Corbett AH (2011) 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* 286(43): 37429-37445.

Fatica A, Dlakic M, Tollervey D (2002) Naf1 p is a box H/ACA snoRNP assembly factor. *RNA* 8(12): 1502-1514.

Fatica A, Morlando M, Bozzoni I (2000) Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. *Embo J* 19(22): 6218-6229.

Ferreira-Cerca S, Poll G, Kuhn H, Neueder A, Jakob S, Tschochner H, Milkereit P (2007) Analysis of the in vivo assembly pathway of eukaryotic 40S ribosomal proteins. *Mol Cell* 28(3): 446-457.

Fong N, Bentley DL (2001) Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* 15(14): 1783-1795.

Fong N, Brannan K, Erickson B, Kim H, Cortazar MA, Sheridan RM, Nguyen T, Karp S, Bentley DL (2015) Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell* 60(2): 256-267.

Fox MJ, Gao H, Smith-Kinnaman WR, Liu Y, Mosley AL (2015) The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway. *PLoS Genet* 11(2): e1004999.

Fronz K, Guttinger S, Burkert K, Kuhn U, Stohr N, Schierhorn A, Wahle E (2011) Arginine methylation of the nuclear poly(a) binding protein weakens the interaction with its nuclear import receptor, transportin. *J Biol Chem* 286(38): 32986-32994.

Fronz K, Otto S, Kolbel K, Kuhn U, Friedrich H, Schierhorn A, Beck-Sickinger AG, Ostareck-Lederer A, Wahle E (2008) Promiscuous modification of the nuclear poly(A)-binding protein by multiple protein-arginine methyltransferases does not affect the aggregation behavior. *J Biol Chem* 283(29): 20408-20420.

Gaillard H, Garcia-Muse T, Aguilera A (2015) Replication stress and cancer. *Nat Rev Cancer* 15(5): 276-289.

Ganot P, Bortolin ML, Kiss T (1997) Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* 89(5): 799-809.

Ganot P, Caizergues-Ferrer M, Kiss T (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev* 11(7): 941-956.

Ganot P, Jady BE, Bortolin ML, Darzacq X, Kiss T (1999) Nucleolar factors direct the 2'-O-ribose methylation and pseudouridylation of U6 spliceosomal RNA. *Mol Cell Biol* 19(10): 6906-6917.

Garas M, Dichtl B, Keller W (2008) The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis. *Rna* 14(12): 2671-2684.

Garcia-Lopez MC, Navarro F (2011) RNA polymerase II conserved protein domains as platforms for protein-protein interactions. *Transcription* 2(4): 193-197.

Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K (2014) Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* 156(4): 812-824.

Gilmartin GM, Fleming ES, Oetjen J, Graveley BR (1995) CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. *Genes Dev* 9(1): 72-83.

Gilmartin GM, Nevins JR (1991) Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of the pre-mRNA. *Mol Cell Biol* 11(5): 2432-2438.

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

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

Graber JH, Cantor CR, Mohr SC, Smith TF (1999) Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic Acids Res* 27(3): 888-894.

Granneman S, Kudla G, Petfalski E, Tollervy D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc Natl Acad Sci U S A* 106(24): 9613-9618.

Grenier St-Sauveur V, Soucek S, Corbett AH, Bachand F (2013) 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* 33(23): 4718-4731.

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

Grosso AR, de Almeida SF, Braga J, Carmo-Fonseca M (2012) Dynamic transitions in RNA polymerase II density profiles during transcription termination. *Genome Res* 22(8): 1447-1456.

Grunberg S, Hahn S (2013) Structural insights into transcription initiation by RNA polymerase II. *Trends Biochem Sci* 38(12): 603-611.

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

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

Gudipati RK, Villa T, Boulay J, Libri D (2008) Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat Struct Mol Biol* 15(8): 786-794.

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

Gullerova M, Proudfoot NJ (2008) Cohesin complex promotes transcriptional termination between convergent genes in *S. pombe*. *Cell* 132(6): 983-995.

Guo Z, Sherman F (1995) 3'-end-forming signals of yeast mRNA. *Mol Cell Biol* 15(11): 5983-5990.

Gupta I, Clauder-Munster S, Klaus B, Jarvelin AI, Aiyar RS, Benes V, Wilkening S, Huber W, Pelechano V, Steinmetz LM (2014) Alternative polyadenylation diversifies post-transcriptional regulation by selective RNA-protein interactions. *Mol Syst Biol* 10(719).

Gusarov I, Nudler E (2001) Control of intrinsic transcription termination by N and NusA: the basic mechanisms. *Cell* 107(4): 437-449.

Hackmann A, Wu H, Schneider UM, Meyer K, Jung K, Krebber H (2014) Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. *Nat Commun* 5(3123).

Haimovich G, Medina DA, Causse SZ, Garber M, Millan-Zambrano G, Barkai O, Chavez S, Perez-Ortin JE, Darzacq X, Choder M (2013) Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* 153(5): 1000-1011.

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

Hammell CM, Gross S, Zenklusen D, Heath CV, Stutz F, Moore C, Cole CN (2002) Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol* 22(18): 6441-6457.

Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, McKnight SL (2012) Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* 149(4): 768-779.

Hardwick SW, Luisi BF (2013) Rarely at rest: RNA helicases and their busy contributions to RNA degradation, regulation and quality control. *RNA Biol* 10(1): 56-70.

Harigaya Y, Tanaka H, Yamanaka S, Tanaka K, Watanabe Y, Tsutsumi C, Chikashige Y, Hiraoka Y, Yamashita A, Yamamoto M (2006) Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature* 442(7098): 45-50.

Hasan A, Cotobal C, Duncan CD, Mata J (2014) Systematic analysis of the role of RNA-binding proteins in the regulation of RNA stability. *PLoS Genet* 10(11): e1004684.

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

Heidemann M, Hintermair C, Voss K, Eick D (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* 1829(1): 55-62.

Henras A, Henry Y, Bousquet-Antonelli C, Noaillac-Depeyre J, Gelugne JP, Caizergues-Ferrer M (1998) Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J* 17(23): 7078-7090.

Henriques T, Gilchrist DA, Nechaev S, Bern M, Muse GW, Burkholder A, Fargo DC, Adelman K (2013) Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol Cell* 52(4): 517-528.

Heo DH, Yoo I, Kong J, Lidschreiber M, Mayer A, Choi BY, Hahn Y, Cramer P, Buratowski S, Kim M (2013) 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* 288(51): 36676-36690.

Hessle V, Bjork P, Sokolowski M, Gonzalez de Valdivia E, Silverstein R, Artemenko K, Tyagi A, Maddalo G, Ilag L, Helbig R, Zubarev RA, Visa N (2009) The exosome associates cotranscriptionally with the nascent pre-mRNP through interactions with heterogeneous nuclear ribonucleoproteins. *Mol Biol Cell* 20(15): 3459-3470.

Hieronimus H, Yu MC, Silver PA (2004) Genome-wide mRNA surveillance is coupled to mRNA export. *Genes Dev* 18(21): 2652-2662.

Hintermair C, Heidemann M, Koch F, Descostes N, Gut M, Gut I, Fenouil R, Ferrier P, Flatley A, Kremmer E, Chapman RD, Andrau JC, Eick D (2012) Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. *EMBO J* 31(12): 2784-2797.

Hirose T, Shu MD, Steitz JA (2003) Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol Cell* 12(1): 113-123.

Hirose Y, Manley JL (1998) RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* 395(6697): 93-96.

Hirose Y, Manley JL (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev* 14(12): 1415-1429.

Ho CK, Shuman S (1999) Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* 3(3): 405-411.

Hobor F, Pergoli R, Kubicek K, Hrossova D, Bacikova V, Zimmermann M, Pasulka J, Hofr C, Vanacova S, Stefl R (2011) Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem* 286(5): 3645-3657.

Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ (2012) RNA polymerase II collision interrupts convergent transcription. *Mol Cell* 48(3): 365-374.

Hoffman CS, Wood V, Fantes PA (2015) An Ancient Yeast for Young Geneticists: A Primer on the *Schizosaccharomyces pombe* Model System. *Genetics* 201(2): 403-423.

Holub P, Lalakova J, Cerna H, Pasulka J, Sarazova M, Hrazdilova K, Arce MS, Hobor F, Stefl R, Vanacova S (2012) Air2p is critical for the assembly and RNA-binding of the TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation. *Nucleic Acids Res* 40(12): 5679-5693.

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

Houseley J, LaCava J, Tollervey D (2006) RNA-quality control by the exosome. *Nat Rev Mol Cell Biol* 7(7): 529-539.

Houseley J, Tollervey D (2006) Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep* 7(2): 205-211.

Hrossova D, Sikorsky T, Potesil D, Bartosovic M, Pasulka J, Zdrahal Z, Stefl R, Vanacova S (2015) RBM7 subunit of the NEXT complex binds U-rich sequences and targets 3'-end extended forms of snRNAs. *Nucleic Acids Res* 43(8): 4236-4248.

Hsin JP, Manley JL (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev* 26(19): 2119-2137.

Hsin JP, Sheth A, Manley JL (2011) RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. *Science* 334(6056): 683-686.

Hu J, Lutz CS, Wilusz J, Tian B (2005) Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA* 11(10): 1485-1493.
Huber W, Toedling J, Steinmetz LM (2006) Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* 22(16): 1963-1970.

Humphrey T, Birse CE, Proudfoot NJ (1994) RNA 3' end signals of the *S.pombe* *ura4* gene comprise a site determining and efficiency element. *EMBO J* 13(10): 2441-2451.

Humphrey T, Sadhale P, Platt T, Proudfoot N (1991) Homologous mRNA 3' end formation in fission and budding yeast. *EMBO J* 10(11): 3503-3511.

Ibrahim H, Wilusz J, Wilusz CJ (2008) RNA recognition by 3'-to-5' exonucleases: the substrate perspective. *Biochim Biophys Acta* 1779(4): 256-265.

Ishii R, Nureki O, Yokoyama S (2003) Crystal structure of the tRNA processing enzyme RNase PH from *Aquifex aeolicus*. *J Biol Chem* 278(34): 32397-32404.

Izban MG, Luse DS (1992) The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'----5' direction in the presence of elongation factor SII. *Genes Dev* 6(7): 1342-1356.

Januszyk K, Lima CD (2011) Structural components and architectures of RNA exosomes. *Adv Exp Med Biol* 702(9-28).

Januszyk K, Lima CD (2014) The eukaryotic RNA exosome. *Curr Opin Struct Biol* 24(132-140).

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

Jeffares DC, Rallis C, Rieux A, Speed D, Prevorovsky M, Mourier T, Marsellach FX, Iqbal Z, Lau W, Cheng TM, Pracana R, Mulleder M, Lawson JL, Chessel A, Bala S, Hellenthal G, O'Fallon B, Keane T, Simpson JT, Bischof L, Tomiczek B, Bitton DA, Sideri T, Codlin S, Hellberg JE, van Trigt L, Jeffery L, Li JJ, Atkinson S, Thodberg M, Febrer M, McLay K, Drou N, Brown W, Hayles J, Carazo Salas RE, Ralser M, Maniatis N, Balding DJ, Balloux F, Durbin R, Bahler J (2015) The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nat Genet* 47(3): 235-241.

Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kuhn U, Menzies FM, Oude Vrielink JA, Bos AJ, Drost J, Rooijers K, Rubinsztein DC, Agami R (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell* 149(3): 538-553.

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

Jeronimo C, Bataille AR, Robert F (2013) The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chem Rev* 113(11): 8491-8522.

Jia H, Wang X, Anderson JT, Jankowsky E (2012) RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proc Natl Acad Sci U S A* 109(19): 7292-7297.

Jia H, Wang X, Liu F, Guenther UP, Srinivasan S, Anderson JT, Jankowsky E (2011) The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. *Cell* 145(6): 890-901.

Jiao Y, Riechmann JL, Meyerowitz EM (2008) Transcriptome-wide analysis of uncapped mRNAs in Arabidopsis reveals regulation of mRNA degradation. *Plant Cell* 20(10): 2571-2585.

Jimeno-Gonzalez S, Schmid M, Malagon F, Haaning LL, Jensen TH (2014) Rat1p maintains RNA polymerase II CTD phosphorylation balance. *RNA* 20(4): 551-558.

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

Kaplan CD, Larsson KM, Kornberg RD (2008) The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell* 30(5): 547-556.

Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149(4): 753-767.

Kaufmann I, Martin G, Friedlein A, Langen H, Keller W (2004) Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. *EMBO J* 23(3): 616-626.

Keller C, Woolcock K, Hess D, Buhler M (2010) Proteomic and functional analysis of the noncanonical poly(A) polymerase Cid14. *RNA* 16(6): 1124-1129.

Keller RW, Kuhn U, Aragon M, Bornikova L, Wahle E, Bear DG (2000) The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. *J Mol Biol* 297(3): 569-583.

Keller W, Bienroth S, Lang KM, Christofori G (1991) Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA. *EMBO J* 10(13): 4241-4249.

Kerwitz Y, Kuhn U, Lilie H, Knoth A, Scheuermann T, Friedrich H, Schwarz E, Wahle E (2003) Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *Embo J* 22(14): 3705-3714.

Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, Moore CL (1997) Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes Dev* 11(19): 2545-2556.

Kierzek E, Malgowska M, Lisowiec J, Turner DH, Gdaniec Z, Kierzek R (2014) The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res* 42(5): 3492-3501.

Kilchert C, Wittmann S, Passoni M, Shah S, Granneman S, Vasiljeva L (2015) Regulation of mRNA Levels by Decay-Promoting Introns that Recruit the Exosome Specificity Factor Mmi1. *Cell Rep* 13(11): 2504-2515.

Kim DU, Hayles J, Kim D, Wood V, Park HO, Won M, Yoo HS, Duhig T, Nam M, Palmer G, Han S, Jeffery L, Baek ST, Lee H, Shim YS, Lee M, Kim L, Heo KS, Noh EJ, Lee AR, Jang YJ, Chung KS, Choi SJ, Park JY, Park Y, Kim HM, Park SK, Park HJ, Kang EJ, Kim HB, Kang HS, Park HM, Kim K, Song K, Song KB, Nurse P, Hoe KL (2010) Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol* 28(6): 617-623.

Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC, Bentley DL (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* 17(10): 1279-1286.

Kim HD, Choe J, Seo YS (1999) The *sen1(+)* gene of *Schizosaccharomyces pombe*, a homologue of budding yeast *SEN1*, encodes an RNA and DNA helicase. *Biochemistry* 38(44): 14697-14710.

Kim K, Heo DH, Kim I, Suh JY, Kim M (2016) Exosome Cofactors Connect Transcription Termination to RNA Processing by Guiding Terminated Transcripts to the Appropriate Exonuclease within the Nuclear Exosome. *J Biol Chem*

Kim M, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S (2004) Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J* 23(2): 354-364.

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

- Kim M, Vasiljeva L, Rando OJ, Zhelkovsky A, Moore C, Buratowski S (2006) Distinct pathways for snoRNA and mRNA termination. *Mol Cell* 24(5): 723-734.
- King TH, Decatur WA, Bertrand E, Maxwell ES, Fournier MJ (2001) A well-connected and conserved nucleoplasmic helicase is required for production of box C/D and H/ACA snoRNAs and localization of snoRNP proteins. *Mol Cell Biol* 21(22): 7731-7746.
- King TH, Liu B, McCully RR, Fournier MJ (2003) Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell* 11(2): 425-435.
- Kinoshita N, Goebel M, Yanagida M (1991) The fission yeast *dis3+* gene encodes a 110-kDa essential protein implicated in mitotic control. *Mol Cell Biol* 11(12): 5839-5847.
- Kishore S, Gruber AR, Jedlinski DJ, Syed AP, Jorjani H, Zavolan M (2013) Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol* 14(5): R45.
- Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T (1996) Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* 85(7): 1077-1088.
- Kiss-Laszlo Z, Henry Y, Kiss T (1998) Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J* 17(3): 797-807.
- Kiss T, Fayet E, Jady BE, Richard P, Weber M (2006) Biogenesis and intranuclear trafficking of human box C/D and H/ACA RNPs. *Cold Spring Harb Symp Quant Biol* 71(407-417).
- Komarnitsky P, Cho EJ, Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* 14(19): 2452-2460.
- Komili S, Silver PA (2008) Coupling and coordination in gene expression processes: a systems biology view. *Nat Rev Genet* 9(1): 38-48.
- Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F, Hofr C, Vanacova S, Stefl R (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26(17): 1891-1896.
- Kuhn U, Gundel M, Knoth A, Kerwitz Y, Rudel S, Wahle E (2009) Poly(A) tail length is controlled by the nuclear poly(A)-binding protein regulating the interaction between poly(A) polymerase and the cleavage and polyadenylation specificity factor. *J Biol Chem* 284(34): 22803-22814.
- Kuhn U, Nemeth A, Meyer S, Wahle E (2003) The RNA binding domains of the nuclear poly(A)-binding protein. *J Biol Chem* 278(19): 16916-16925.

Kuhn U, Wahle E (2004) Structure and function of poly(A) binding proteins. *Biochim Biophys Acta* 1678(2-3): 67-84.

Kulish D, Struhl K (2001) TFIIS enhances transcriptional elongation through an artificial arrest site in vivo. *Mol Cell Biol* 21(13): 4162-4168.

Kwok CK, Tang Y, Assmann SM, Bevilacqua PC (2015) The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends Biochem Sci* 40(4): 221-232.

Kwon I, Kato M, Xiang S, Wu L, Theodoropoulos P, Mirzaei H, Han T, Xie S, Corden JL, McKnight SL (2013) Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. *Cell* 155(5): 1049-1060.

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

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

Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi EA, Xie X, Wan J, Xing Y, Freudenberg JM, Yang P, Jothi R, Hu G, Shi Y (2014) Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J* 33(8): 878-889.

Lackner DH, Beilharz TH, Marguerat S, Mata J, Watt S, Schubert F, Preiss T, Bahler J (2007) A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol Cell* 26(1): 145-155.

Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D (1998) The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes Dev* 12(4): 527-537.

Lafontaine DL, Tollervey D (1999) Nop58p is a common component of the box C+D snoRNPs that is required for snoRNA stability. *RNA* 5(3): 455-467.

Lai D, Proctor JR, Meyer IM (2013) On the importance of cotranscriptional RNA structure formation. *RNA* 19(11): 1461-1473.

Larochelle M, Lemay JF, Bachand F (2012) The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. *Nucleic Acids Res* 40(20): 10240-10253.

Lebreton A, Seraphin B (2008) Exosome-mediated quality control: substrate recruitment and molecular activity. *Biochim Biophys Acta* 1779(9): 558-565.

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

Lee NN, Chalamcharla VR, Reyes-Turcu F, Mehta S, Zofall M, Balachandran V, Dhakshnamoorthy J, Taneja N, Yamanaka S, Zhou M, Grewal SI (2013) Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. *Cell* 155(5): 1061-1074.

Legros P, Malapert A, Niinuma S, Bernard P, Vanoosthuyse V (2014) RNA processing factors Swd2.2 and Sen1 antagonize RNA Pol III-dependent transcription and the localization of condensin at Pol III genes. *PLoS Genet* 10(11): e1004794.

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

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

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

Lemieux C, Bachand F (2009) Cotranscriptional recruitment of the nuclear poly(A)-binding protein Pab2 to nascent transcripts and association with translating mRNPs. *Nucleic Acids Res* 37(10): 3418-3430.

Lemieux C, Marguerat S, Lafontaine J, Barbezier N, Bahler J, Bachand F (2011) A Pre-mRNA degradation pathway that selectively targets intron-containing genes requires the nuclear poly(A)-binding protein. *Mol Cell* 44(1): 108-119.

Lenstra TL, Tudek A, Clauder S, Xu Z, Pachis ST, van Leenen D, Kemmeren P, Steinmetz LM, Libri D, Holstege FC (2013) The role of Ctk1 kinase in termination of small non-coding RNAs. *PLoS One* 8(12): e80495.

Li W, You B, Hoque M, Zheng D, Luo W, Ji Z, Park JY, Gunderson SI, Kalsotra A, Manley JL, Tian B (2015) 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* 11(4): e1005166.

Li Z, Pandit S, Deutscher MP (1998) Polyadenylation of stable RNA precursors in vivo. *Proc Natl Acad Sci U S A* 95(21): 12158-12162.

Libri D (2015) Endless Quarrels at the End of Genes. *Mol Cell* 60(2): 192-194.

- Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* 9(5): 1101-1111.
- Lim SJ, Boyle PJ, Chinen M, Dale RK, Lei EP (2013) Genome-wide localization of exosome components to active promoters and chromatin insulators in *Drosophila*. *Nucleic Acids Res* 41(5): 2963-2980.
- Lingner J, Radtke I, Wahle E, Keller W (1991) Purification and characterization of poly(A) polymerase from *Saccharomyces cerevisiae*. *J Biol Chem* 266(14): 8741-8746.
- Liu JJ, Bratkowski MA, Liu X, Niu CY, Ke A, Wang HW (2014) Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM. *Nat Struct Mol Biol* 21(1): 95-102.
- Liu P, Kenney JM, Stiller JW, Greenleaf AL (2010) Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol Biol Evol* 27(11): 2628-2641.
- Liu Q, Greimann JC, Lima CD (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* 127(6): 1223-1237.
- Logan J, Falck-Pedersen E, Darnell JE, Jr., Shenk T (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A* 84(23): 8306-8310.
- Lorentzen E, Basquin J, Tomecki R, Dziembowski A, Conti E (2008) Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol Cell* 29(6): 717-728.
- Loya TJ, O'Rourke TW, Reines D (2013) Yeast Nab3 protein contains a self-assembly domain found in human heterogeneous nuclear ribonucleoprotein-C (hnRNP-C) that is necessary for transcription termination. *J Biol Chem* 288(4): 2111-2117.
- Lubas M, Andersen PR, Schein A, Dziembowski A, Kudla G, Jensen TH (2015) The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* 10(2): 178-192.
- Lubas M, Christensen MS, Kristiansen MS, Domanski M, Falkenby LG, Lykke-Andersen S, Andersen JS, Dziembowski A, Jensen TH (2011) Interaction profiling identifies the human nuclear exosome targeting complex. *Mol Cell* 43(4): 624-637.
- Lunde BM, Reichow SL, Kim M, Suh H, Leeper TC, Yang F, Mutschler H, Buratowski S, Meinhart A, Varani G (2010) Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17(10): 1195-1201.

- Luo W, Johnson AW, Bentley DL (2006) The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev* 20(8): 954-965.
- Ma X, Yang C, Alexandrov A, Grayhack EJ, Behm-Ansmant I, Yu YT (2005) Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J* 24(13): 2403-2413.
- MacDonald CC, Wilusz J, Shenk T (1994) The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. *Mol Cell Biol* 14(10): 6647-6654.
- Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, Helm M, Bujnicki JM, Grosjean H (2013) MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res* 41(Database issue): D262-267.
- Makino DL, Baumgartner M, Conti E (2013) Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. *Nature* 495(7439): 70-75.
- Makino DL, Schuch B, Stegmann E, Baumgartner M, Basquin C, Conti E (2015) RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature* 524(7563): 54-58.
- Mallet PL, Bachand F (2013) A proline-tyrosine nuclear localization signal (PY-NLS) is required for the nuclear import of fission yeast PAB2, but not of human PABPN1. *Traffic* 14(3): 282-294.
- Mandel CR, Bai Y, Tong L (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* 65(7-8): 1099-1122.
- Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* 444(7121): 953-956.
- Mangus DA, Evans MC, Jacobson A (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* 4(7): 223.
- Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R, Bahler J (2012) Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* 151(3): 671-683.
- Marina DB, Shankar S, Natarajan P, Finn KJ, Madhani HD (2013) A conserved ncRNA-binding protein recruits silencing factors to heterochromatin through an RNAi-independent mechanism. *Genes Dev* 27(17): 1851-1856.
- Marquardt S, Hazelbaker DZ, Buratowski S (2011) Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription* 2(3): 145-154.

Martin-Tumasch S, Brow DA (2015) *Saccharomyces cerevisiae* Sen1 Helicase Domain Exhibits 5'- to 3'-Helicase Activity with a Preference for Translocation on DNA Rather than RNA. *J Biol Chem* 290(38): 22880-22889.

Martin G, Gruber AR, Keller W, Zavolan M (2012) Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep* 1(6): 753-763.

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

Masuda A, Takeda J, Okuno T, Okamoto T, Ohkawara B, Ito M, Ishigaki S, Sobue G, Ohno K (2015) Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes Dev* 29(10): 1045-1057.

Mata J (2013) Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol* 10(8): 1407-1414.

Matsuyama A, Arai R, Yashiroda Y, Shirai A, Kamata A, Sekido S, Kobayashi Y, Hashimoto A, Hamamoto M, Hiraoka Y, Horinouchi S, Yoshida M (2006) ORFeome cloning and global analysis of protein localization in the fission yeast *Schizosaccharomyces pombe*. *Nat Biotechnol* 24(7): 841-847.

Maundrell K (1993) Thiamine-repressible expression vectors pREP and pRIP for fission yeast. *Gene* 123(1): 127-130.

Mayer A, Heidemann M, Lidschreiber M, Schrieck A, Sun M, Hintermair C, Kremmer E, Eick D, Cramer P (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336(6089): 1723-1725.

Mayer A, Lidschreiber M, Siebert M, Leike K, Soding J, Cramer P (2010) Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* 17(10): 1272-1278.

Mayer A, Schrieck A, Lidschreiber M, Leike K, Martin DE, Cramer P (2012) The spt5 C-terminal region recruits yeast 3' RNA cleavage factor I. *Mol Cell Biol* 32(7): 1321-1331.

Mayr C (2016) Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol* 26(3): 227-237.

Mayr C, Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138(4): 673-684.

McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL (1997) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* 11(24): 3306-3318.

McCracken S, Fong N, Yankulov K, Ballantyne S, Pan G, Greenblatt J, Patterson SD, Wickens M, Bentley DL (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* 385(6614): 357-361.

McMahon M, Contreras A, Ruggero D (2015) Small RNAs with big implications: new insights into H/ACA snoRNA function and their role in human disease. *Wiley Interdiscip Rev RNA* 6(2): 173-189.

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

Meyer S, Urbanke C, Wahle E (2002) Equilibrium studies on the association of the nuclear poly(A) binding protein with poly(A) of different lengths. *Biochemistry* 41(19): 6082-6089.
Midtgaard SF, Assenholt J, Jonstrup AT, Van LB, Jensen TH, Brodersen DE (2006) 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* 103(32): 11898-11903.

Milligan L, Torchet C, Allmang C, Shipman T, Tollervey D (2005) A nuclear surveillance pathway for mRNAs with defective polyadenylation. *Mol Cell Biol* 25(22): 9996-10004.

Mischo HE, Proudfoot NJ (2013) Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta* 1829(1): 174-185.

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

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

Mitsuzawa H, Kanda E, Ishihama A (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res* 31(16): 4696-4701.

Mohanty BK, Kushner SR (2006) The majority of Escherichia coli mRNAs undergo post-transcriptional modification in exponentially growing cells. *Nucleic Acids Res* 34(19): 5695-5704.

Mojardin L, Vazquez E, Antequera F (2013) Specification of DNA replication origins and genomic base composition in fission yeasts. *J Mol Biol* 425(23): 4706-4713.

Moreira MC, Klur S, Watanabe M, Nemeth AH, Le Ber I, Moniz JC, Tranchant C, Aubourg P, Tazir M, Schols L, Pandolfo M, Schulz JB, Pouget J, Calvas P, Shizuka-Ikeda M, Shoji M, Tanaka M, Izatt L, Shaw CE, M'Zahem A, Dunne E, Bomont P, Benhassine T, Bouslam N, Stevanin G, Brice A, Guimaraes J, Mendonca P, Barbot C, Coutinho P, Sequeiros J, Durr

A, Warter JM, Koenig M (2004) Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat Genet* 36(3): 225-227.

Morlando M, Ballarino M, Greco P, Caffarelli E, Dichtl B, Bozzoni I (2004) Coupling between snoRNP assembly and 3' processing controls box C/D snoRNA biosynthesis in yeast. *EMBO J* 23(12): 2392-2401.

Motorin Y, Helm M (2011) RNA nucleotide methylation. *Wiley Interdiscip Rev RNA* 2(5): 611-631.

Moyle-Heyrman G, Zaichuk T, Xi L, Zhang Q, Uhlenbeck OC, Holmgren R, Widom J, Wang JP (2013) Chemical map of *Schizosaccharomyces pombe* reveals species-specific features in nucleosome positioning. *Proc Natl Acad Sci U S A* 110(50): 20158-20163.

Muhlbacher W, Mayer A, Sun M, Remmert M, Cheung AC, Niesser J, Soeding J, Cramer P (2015) Structure of Ctk3, a subunit of the RNA polymerase II CTD kinase complex, reveals a noncanonical CTD-interacting domain fold. *Proteins* 83(10): 1849-1858.

Muller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, Curk T, Poser I, Zarnack K, Neugebauer KM (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev* 30(5): 553-566.

Murakami H, Goto DB, Toda T, Chen ES, Grewal SI, Martienssen RA, Yanagida M (2007) Ribonuclease activity of Dis3 is required for mitotic progression and provides a possible link between heterochromatin and kinetochore function. *PLoS ONE* 2(3): e317.

Murthy KG, Manley JL (1992) Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J Biol Chem* 267(21): 14804-14811.

Murthy KG, Manley JL (1995) The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev* 9(21): 2672-2683.

Nag A, Narsinh K, Kazerouninia A, Martinson HG (2006) The conserved AAUAAA hexamer of the poly(A) signal can act alone to trigger a stable decrease in RNA polymerase II transcription velocity. *RNA* 12(8): 1534-1544.

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

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

Nemeth A, Krause S, Blank D, Jenny A, Jenó P, Lustig A, Wahle E (1995) Isolation of genomic and cDNA clones encoding bovine poly(A) binding protein II. *Nucleic Acids Res* 23(20): 4034-4041.

Nguyen D, Grenier St-Sauveur V, Bergeron D, Dupuis-Sandoval F, Scott MS, Bachand F (2015) A Polyadenylation-Dependent 3' End Maturation Pathway Is Required for the Synthesis of the Human Telomerase RNA. *Cell Rep* 13(10): 2244-2257.

Nicastro G, Taylor IA, Ramos A (2015) KH-RNA interactions: back in the groove. *Curr Opin Struct Biol* 30(63-70).

Nicoloso M, Qu LH, Michot B, Bachellerie JP (1996) Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs. *J Mol Biol* 260(2): 178-195.

Nielsen S, Yuzenkova Y, Zenkin N (2013) Mechanism of eukaryotic RNA polymerase III transcription termination. *Science* 340(6140): 1577-1580.

Nojima T, Gomes T, Grosso AR, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ (2015) Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* 161(3): 526-540.

Ntini E, Jarvelin AI, Bornholdt J, Chen Y, Boyd M, Jorgensen M, Andersson R, Hoof I, Schein A, Andersen PR, Andersen PK, Preker P, Valen E, Zhao X, Pelechano V, Steinmetz LM, Sandelin A, Jensen TH (2013) Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nat Struct Mol Biol* 20(8): 923-928.

Nudler E (2012) RNA polymerase backtracking in gene regulation and genome instability. *Cell* 149(7): 1438-1445.

O'Hara EB, Chekanova JA, Ingle CA, Kushner ZR, Peters E, Kushner SR (1995) Polyadenylation helps regulate mRNA decay in *Escherichia coli*. *Proc Natl Acad Sci U S A* 92(6): 1807-1811.

O'Reilly D, Kuznetsova OV, Laitem C, Zaborowska J, Dienstbier M, Murphy S (2014) Human snRNA genes use polyadenylation factors to promote efficient transcription termination. *Nucleic Acids Res* 42(1): 264-275.

O'Rourke TW, Loya TJ, Head PE, Horton JR, Reines D (2015) Amyloid-like assembly of the low complexity domain of yeast Nab3. *Prion* 9(1): 34-47.

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

Orozco IJ, Kim SJ, Martinson HG (2002) The poly(A) signal, without the assistance of any downstream element, directs RNA polymerase II to pause in vivo and then to release stochastically from the template. *J Biol Chem* 277(45): 42899-42911.

Osheim YN, Proudfoot NJ, Beyer AL (1999) EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol Cell* 3(3): 379-387.

Osheim YN, Sikes ML, Beyer AL (2002) EM visualization of Pol II genes in *Drosophila*: most genes terminate without prior 3' end cleavage of nascent transcripts. *Chromosoma* 111(1): 1-12.

Pan X, Lei B, Zhou N, Feng B, Yao W, Zhao X, Yu Y, Lu H (2012) Identification of novel genes involved in DNA damage response by screening a genome-wide *Schizosaccharomyces pombe* deletion library. *BMC Genomics* 13(662).

Pearson EL, Moore CL (2013) Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor Rat1. *J Biol Chem* 288(27): 19750-19759.

Perreault A, Bellemer C, Bachand F (2008) Nuclear export competence of pre-40S subunits in fission yeast requires the ribosomal protein Rps2. *Nucleic Acids Res* 36(19): 6132-6142.

Perreault A, Lemieux C, Bachand F (2007) Regulation of the Nuclear Poly(A)-binding Protein by Arginine Methylation in Fission Yeast. *J Biol Chem* 282(10): 7552-7562.

Peters JM, Vangeloff AD, Landick R (2011) Bacterial transcription terminators: the RNA 3'-end chronicles. *J Mol Biol* 412(5): 793-813.

Piekna-Przybylska D, Decatur WA, Fournier MJ (2007) New bioinformatic tools for analysis of nucleotide modifications in eukaryotic rRNA. *RNA* 13(3): 305-312.

Pinto PA, Henriques T, Freitas MO, Martins T, Domingues RG, Wyrzykowska PS, Coelho PA, Carmo AM, Sunkel CE, Proudfoot NJ, Moreira A (2011) RNA polymerase II kinetics in polo polyadenylation signal selection. *EMBO J* 30(12): 2431-2444.

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

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

Porrúa O, Libri D (2013) RNA quality control in the nucleus: the Angels' share of RNA. *Biochim Biophys Acta* 1829(6-7): 604-611.

Porrua O, Libri D (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* 16(3): 190-202.

Preiss T, Hentze MW (1998) Dual function of the messenger RNA cap structure in poly(A)-tail-promoted translation in yeast. *Nature* 392(6675): 516-520.

Preker P, Almvig K, Christensen MS, Valen E, Mapendano CK, Sandelin A, Jensen TH (2011) PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Res* 39(16): 7179-7193.

Preker PJ, Lingner J, Minvielle-Sebastia L, Keller W (1995) The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. *Cell* 81(3): 379-389.

Prescott EM, Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A* 99(13): 8796-8801.

Proudfoot N (2004) New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* 16(3): 272-278.

Qiu H, Hu C, Hinnebusch AG (2009) Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell* 33(6): 752-762.

Raabe T, Bollum FJ, Manley JL (1991) Primary structure and expression of bovine poly(A) polymerase. *Nature* 353(6341): 229-234.

Rammelt C, Bilen B, Zavolan M, Keller W (2011) PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif. *RNA* 17(9): 1737-1746.

Rasmussen TP, Culbertson MR (1998) The putative nucleic acid helicase Sen1p is required for formation and stability of termini and for maximal rates of synthesis and levels of accumulation of small nucleolar RNAs in *Saccharomyces cerevisiae*. *Mol Cell Biol* 18(12): 6885-6896.

Reichow SL, Hama T, Ferre-D'Amare AR, Varani G (2007) The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res* 35(5): 1452-1464.

Reines D, Ghanouni P, Li QQ, Mote J, Jr. (1992) The RNA polymerase II elongation complex. Factor-dependent transcription elongation involves nascent RNA cleavage. *J Biol Chem* 267(22): 15516-15522.

Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF, Heiman DI, Young SK, Furuya K, Guo Y, Pidoux A, Chen HM, Robbertse B, Goldberg JM, Aoki K, Bayne EH, Berlin AM, Desjardins CA, Dobbs E, Dukaj L, Fan L, FitzGerald MG, French C, Gujja S, Hansen K, Keifenheim D, Levin JZ, Mosher RA, Muller CA, Pfiffner

J, Priest M, Russ C, Smialowska A, Swoboda P, Sykes SM, Vaughn M, Vengrova S, Yoder R, Zeng Q, Allshire R, Baulcombe D, Birren BW, Brown W, Ekwall K, Kellis M, Leatherwood J, Levin H, Margalit H, Martienssen R, Nieduszynski CA, Spatafora JW, Friedman N, Dalggaard JZ, Baumann P, Niki H, Regev A, Nusbaum C (2011) Comparative functional genomics of the fission yeasts. *Science* 332(6032): 930-936.

Richard P, Kiss AM, Darzacq X, Kiss T (2006) Cotranscriptional recognition of human intronic box H/ACA snoRNAs occurs in a splicing-independent manner. *Mol Cell Biol* 26(7): 2540-2549.

Robinson SR, Oliver AW, Chevassut TJ, Newbury SF (2015) The 3' to 5' Exoribonuclease DIS3: From Structure and Mechanisms to Biological Functions and Role in Human Disease. *Biomolecules* 5(3): 1515-1539.

Roguev A, Shevchenko A, Schaft D, Thomas H, Stewart AF (2004) A comparative analysis of an orthologous proteomic environment in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Mol Cell Proteomics* 3(2): 125-132.

Romero P, Obradovic Z, Dunker AK (2004) Natively disordered proteins: functions and predictions. *Appl Bioinformatics* 3(2-3): 105-113.

Rosonina E, Yurko N, Li W, Hoque M, Tian B, Manley JL (2014) Threonine-4 of the budding yeast RNAP II CTD couples transcription with Htz1-mediated chromatin remodeling. *Proc Natl Acad Sci U S A* 111(33): 11924-11931.

Rozenblatt-Rosen O, Nagaike T, Francis JM, Kaneko S, Glatt KA, Hughes CM, LaFramboise T, Manley JL, Meyerson M (2009) The tumor suppressor Cdc73 functionally associates with CPSF and CstF 3' mRNA processing factors. *Proc Natl Acad Sci U S A* 106(3): 755-760.

Ruegsegger U, Beyer K, Keller W (1996) Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J Biol Chem* 271(11): 6107-6113.

Russell J, Zomerdijk JC (2005) RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem Sci* 30(2): 87-96.

Russo P, Li WZ, Guo Z, Sherman F (1993) Signals that produce 3' termini in CYC1 mRNA of the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* 13(12): 7836-7849.

Russo P, Li WZ, Hampsey DM, Zaret KS, Sherman F (1991) Distinct cis-acting signals enhance 3' endpoint formation of CYC1 mRNA in the yeast *Saccharomyces cerevisiae*. *EMBO J* 10(3): 563-571.

Ryan K, Calvo O, Manley JL (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* 10(4): 565-573.

Sadowski M, Dichtl B, Hubner W, Keller W (2003) Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J* 22(9): 2167-2177.

Saguez C, Schmid M, Olesen JR, Ghazy MA, Qu X, Poulsen MB, Nasser T, Moore C, Jensen TH (2008) Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. *Mol Cell* 31(1): 91-103.

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

Schaeffer D, Reis FP, Johnson SJ, Arraiano CM, van Hoof A (2012) The CR3 motif of Rrp44p is important for interaction with the core exosome and exosome function. *Nucleic Acids Res* 40(18): 9298-9307.

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

Schaughency P, Merran J, Corden JL (2014) Genome-wide mapping of yeast RNA polymerase II termination. *PLoS Genet* 10(10): e1004632.

Schlackow M, Marguerat S, Proudfoot NJ, Bahler J, Erban R, Gullerova M (2013) Genome-wide analysis of poly(A) site selection in *Schizosaccharomyces pombe*. *RNA* 19(12): 1617-1631.

Schlackow M, Marguerat S, Proudfoot NJ, Bähler J, Erban R, Gullerova M (2013) Genome-wide analysis of poly(A) site selection in *Schizosaccharomyces pombe*. *RNA*

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

Schmid M, Jensen TH (2013) Transcription-associated quality control of mRNP. *Biochim Biophys Acta* 1829(1): 158-168.

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

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

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

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

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

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

Schneider S, Pei Y, Shuman S, Schwer B (2010) Separable functions of the fission yeast Spt5 carboxyl-terminal domain (CTD) in capping enzyme binding and transcription elongation overlap with those of the RNA polymerase II CTD. *Mol Cell Biol* 30(10): 2353-2364.

Schonemann L, Kuhn U, Martin G, Schafer P, Gruber AR, Keller W, Zavolan M, Wahle E (2014) Reconstitution of CPSF active in polyadenylation: recognition of the polyadenylation signal by WDR33. *Genes Dev* 28(21): 2381-2393.

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

Schuller R, Forne I, Straub T, Schrieck A, Texier Y, Shah N, Decker TM, Cramer P, Imhof A, Eick D (2016) Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol Cell* 61(2): 305-314.

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

Schwartz S, Bernstein DA, Mumbach MR, Jovanovic M, Herbst RH, Leon-Ricardo BX, Engreitz JM, Guttman M, Satija R, Lander ES, Fink G, Regev A (2014) Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* 159(1): 148-162.

Schwer B, Shuman S (2011) Deciphering the RNA polymerase II CTD code in fission yeast. *Mol Cell* 43(2): 311-318.

Shah S, Wittmann S, Kilchert C, Vasiljeva L (2014) lncRNA recruits RNAi and the exosome to dynamically regulate *pho1* expression in response to phosphate levels in fission yeast. *Genes Dev* 28(3): 231-244.

Shearwin KE, Callen BP, Egan JB (2005) Transcriptional interference--a crash course. *Trends Genet* 21(6): 339-345.

Sheets MD, Wickens M (1989) Two phases in the addition of a poly(A) tail. *Genes Dev* 3(9): 1401-1412.

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd, Frank J, Manley JL (2009) Molecular architecture of the human pre-mRNA 3' processing complex. *Mol Cell* 33(3): 365-376.

Shi Y, Manley JL (2015) The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes Dev* 29(9): 889-897.

Shi Z, Yang WZ, Lin-Chao S, Chak KF, Yuan HS (2008) Crystal structure of Escherichia coli PNPase: central channel residues are involved in processive RNA degradation. *RNA* 14(11): 2361-2371.

Sigurdsson S, Dirac-Svejstrup AB, Svejstrup JQ (2010) Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell* 38(2): 202-210.

Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH (2009) JBrowse: a next-generation genome browser. *Genome Res* 19(9): 1630-1638.

Skourti-Stathaki K, Kamieniarz-Gdula K, Proudfoot NJ (2014) R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516(7531): 436-439.

Skourti-Stathaki K, Proudfoot NJ (2014) A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* 28(13): 1384-1396.

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

Smith JJ, Rucknagel KP, Schierhorn A, Tang J, Nemeth A, Linder M, Herschman HR, Wahle E (1999) Unusual sites of arginine methylation in Poly(A)-binding protein II and in vitro methylation by protein arginine methyltransferases PRMT1 and PRMT3. *J Biol Chem* 274(19): 13229-13234.

Somesh BP, Reid J, Liu WF, Sogaard TM, Erdjument-Bromage H, Tempst P, Svejstrup JQ (2005) Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell* 121(6): 913-923.

Sonkar A, Yadav S, Ahmed S (2016) Cleavage and polyadenylation factor, Rna14 is an essential protein required for the maintenance of genomic integrity in fission yeast *Schizosaccharomyces pombe*. *Biochim Biophys Acta* 1863(2): 189-197.

Spahr H, Calero G, Bushnell DA, Kornberg RD (2009) *Schizosaccharomyces pombe* RNA polymerase II at 3.6-Å resolution. *Proc Natl Acad Sci U S A* 106(23): 9185-9190.

St-Andre O, Lemieux C, Perreault A, Lackner DH, Bahler J, Bachand F (2010) Negative regulation of meiotic gene expression by the nuclear poly(a)-binding protein in fission yeast. *J Biol Chem* 285(36): 27859-27868.

Stead JA, Costello JL, Livingstone MJ, Mitchell P (2007) 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* 35(16): 5556-5567.

Steinmetz EJ, Brow DA (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol Cell Biol* 16(12): 6993-7003.

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

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

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

Stepanov GA, Filippova JA, Komissarov AB, Kuligina EV, Richter VA, Semenov DV (2015) Regulatory role of small nucleolar RNAs in human diseases. *Biomed Res Int* 2015(206849).

Subtelny AO, Eichhorn SW, Chen GR, Sive H, Bartel DP (2014) Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* 508(7494): 66-71.

Sugiyama T, Sugioka-Sugiyama R (2011) Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast. *EMBO J* 30(6): 1027-1039.

Sugiyama T, Sugioka-Sugiyama R, Hada K, Niwa R (2012) Rhn1, a nuclear protein, is required for suppression of meiotic mRNAs in mitotically dividing fission yeast. *PLoS One* 7(8): e42962.

Sugiyama T, Wanatabe N, Kitahata E, Tani T, Sugioka-Sugiyama R (2013) Red5 and three nuclear pore components are essential for efficient suppression of specific mRNAs during vegetative growth of fission yeast. *Nucleic Acids Res* 41(13): 6674-6686.

Suh H, Ficarro SB, Kang UB, Chun Y, Marto JA, Buratowski S (2016) Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Mol Cell* 61(2): 297-304.

Sun M, Schwalb B, Pirkl N, Maier KC, Schenk A, Failmezger H, Tresch A, Cramer P (2013) Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Mol Cell* 52(1): 52-62.

Szewczak LB, DeGregorio SJ, Strobel SA, Steitz JA (2002) Exclusive interaction of the 15.5 kD protein with the terminal box C/D motif of a methylation guide snoRNP. *Chem Biol* 9(10): 1095-1107.

Takagaki Y, Manley JL (1997) RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol* 17(7): 3907-3914.

Takagaki Y, Manley JL (2000) Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* 20(5): 1515-1525.

Takagaki Y, Manley JL, MacDonald CC, Wilusz J, Shenk T (1990) A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs. *Genes Dev* 4(12A): 2112-2120.

Takagaki Y, Ryner LC, Manley JL (1989) Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev* 3(11): 1711-1724.

Takagaki Y, Seipelt RL, Peterson ML, Manley JL (1996) The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* 87(5): 941-952.

Tardiff DF, Lacadie SA, Rosbash M (2006) A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. *Mol Cell* 24(6): 917-929.

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

Thoms M, Thomson E, Bassler J, Gnadig M, Griesel S, Hurt E (2015) The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell* 162(5): 1029-1038.

Tian B, Graber JH (2012) Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip Rev RNA* 3(3): 385-396.

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

Tietjen JR, Zhang DW, Rodriguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X, Chapman RD, Shokat K, Keles S, Eick D, Ansari AZ (2010) Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* 17(9): 1154-1161.

Tollervey D, Lehtonen H, Carmo-Fonseca M, Hurt EC (1991) The small nucleolar RNP protein NOP1 (fibrillarin) is required for pre-rRNA processing in yeast. *EMBO J* 10(3): 573-583.

Tollervey D, Lehtonen H, Jansen R, Kern H, Hurt EC (1993) Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. *Cell* 72(3): 443-457.

Tomecki R, Drazkowska K, Kucinski I, Stodus K, Szczesny RJ, Gruchota J, Owczarek EP, Kalisiak K, Dziembowski A (2014) Multiple myeloma-associated hDIS3 mutations cause perturbations in cellular RNA metabolism and suggest hDIS3 PIN domain as a potential drug target. *Nucleic Acids Res* 42(2): 1270-1290.

Tomson BN, Crisucci EM, Heisler LE, Gebbia M, Nislow C, Arndt KM (2013) Effects of the Paf1 complex and histone modifications on snoRNA 3'-end formation reveal broad and locus-specific regulation. *Mol Cell Biol* 33(1): 170-182.

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

Toussaint M, Conconi A (2006) High-throughput and sensitive assay to measure yeast cell growth: a bench protocol for testing genotoxic agents. *Nat Protoc* 1(4): 1922-1928.

Trahan C, Aguilar LC, Oeffinger M (2016) Single-Step Affinity Purification (ssAP) and Mass Spectrometry of Macromolecular Complexes in the Yeast *S. cerevisiae*. *Methods Mol Biol* 1361(265-287).

Tuck AC, Tollervey D (2013) A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. *Cell* 154(5): 996-1009.

Tucker JF, Ohle C, Schermann G, Bendrin K, Zhang W, Fischer T, Zhang K (2016) A Novel Epigenetic Silencing Pathway Involving the Highly Conserved 5'-3' Exoribonuclease Dhp1/Rat1/Xrn2 in *Schizosaccharomyces pombe*. *PLoS Genet* 12(2): e1005873.

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

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98(9): 5116-5121.

Tycowski KT, You ZH, Graham PJ, Steitz JA (1998) Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol Cell* 2(5): 629-638.

Ursic D, Chinchilla K, Finkel JS, Culbertson MR (2004) Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res* 32(8): 2441-2452.

- van Hoof A, Lennertz P, Parker R (2000) Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs. *Mol Cell Biol* 20(2): 441-452.
- Vanacova S, Stefl R (2007) The exosome and RNA quality control in the nucleus. *EMBO Rep* 8(7): 651-657.
- Vanacova S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol* 3(6): e189.
- Vanoosthuyse V, Legros P, van der Sar SJ, Yvert G, Toda K, Le Bihan T, Watanabe Y, Hardwick K, Bernard P (2014) CPF-associated phosphatase activity opposes condensin-mediated chromosome condensation. *PLoS Genet* 10(6): e1004415.
- Vasiljeva L, Buratowski S (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol Cell* 21(2): 239-248.
- Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 15(8): 795-804.
- Vasiljeva L, Kim M, Terzi N, Soares LM, Buratowski S (2008) Transcription termination and RNA degradation contribute to silencing of RNA polymerase II transcription within heterochromatin. *Mol Cell* 29(3): 313-323.
- Venkataraman K, Brown KM, Gilmartin GM (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* 19(11): 1315-1327.
- Vo TV, Das J, Meyer MJ, Cordero NA, Akturk N, Wei X, Fair BJ, Degatano AG, Fragoza R, Liu LG, Matsuyama A, Trickey M, Horibata S, Grimson A, Yamano H, Yoshida M, Roth FP, Pleiss JA, Xia Y, Yu H (2016) A Proteome-wide Fission Yeast Interactome Reveals Network Evolution Principles from Yeasts to Human. *Cell* 164(1-2): 310-323.
- von Hippel PH (1998) An integrated model of the transcription complex in elongation, termination, and editing. *Science* 281(5377): 660-665.
- Wagschal A, Rousset E, Basavarajaiah P, Contreras X, Harwig A, Laurent-Chabalier S, Nakamura M, Chen X, Zhang K, Meziane O, Boyer F, Parrinello H, Berkhout B, Terzian C, Benkirane M, Kiernan R (2012) Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* 150(6): 1147-1157.
- Wahle E (1991) A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* 66(4): 759-768.

Wahle E (1991) Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors. *J Biol Chem* 266(5): 3131-3139.

Wahle E (1995) Poly(A) tail length control is caused by termination of processive synthesis. *J Biol Chem* 270(6): 2800-2808.

Wahle E, Martin G, Schiltz E, Keller W (1991) Isolation and expression of cDNA clones encoding mammalian poly(A) polymerase. *EMBO J* 10(13): 4251-4257.

Wang CY, Chen WL, Wang SW (2013) Pdc1 functions in the assembly of P bodies in *Schizosaccharomyces pombe*. *Mol Cell Biol* 33(6): 1244-1253.

Wang HW, Wang J, Ding F, Callahan K, Bratkowski MA, Butler JS, Nogales E, Ke A (2007) Architecture of the yeast Rrp44 exosome complex suggests routes of RNA recruitment for 3' end processing. *Proc Natl Acad Sci U S A* 104(43): 16844-16849.

Wang SW, Asakawa K, Win TZ, Toda T, Norbury CJ (2005) Inactivation of the pre-mRNA cleavage and polyadenylation factor Pfs2 in fission yeast causes lethal cell cycle defects. *Mol Cell Biol* 25(6): 2288-2296.

Wang SW, Stevenson AL, Kearsley SE, Watt S, Bahler J (2008) Global role for polyadenylation-assisted nuclear RNA degradation in posttranscriptional gene silencing. *Mol Cell Biol* 28(2): 656-665.

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

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

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

Watkins NJ, Bohnsack MT (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* 3(3): 397-414.

Watkins NJ, Dickmanns A, Luhrmann R (2002) Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. *Mol Cell Biol* 22(23): 8342-8352.

Watkins NJ, Segault V, Charpentier B, Nottrott S, Fabrizio P, Bachi A, Wilm M, Rosbash M, Branlant C, Luhrmann R (2000) A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP. *Cell* 103(3): 457-466.

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

West ML, Corden JL (1995) Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics* 140(4): 1223-1233.

West S, Gromak N, Norbury CJ, Proudfoot NJ (2006) Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Mol Cell* 21(3): 437-443.

West S, Gromak N, Proudfoot NJ (2004) Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432(7016): 522-525.

West S, Proudfoot NJ (2008) Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic Acids Res* 36(3): 905-914.

Whitelaw E, Proudfoot N (1986) Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J* 5(11): 2915-2922.

Wigington CP, Williams KR, Meers MP, Bassell GJ, Corbett AH (2014) Poly(A) RNA-binding proteins and polyadenosine RNA: new members and novel functions. *Wiley Interdiscip Rev RNA* 5(5): 601-622.

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453(7199): 1239-1243.

Wilson SM, Datar KV, Paddy MR, Swedlow JR, Swanson MS (1994) Characterization of nuclear polyadenylated RNA-binding proteins in *Saccharomyces cerevisiae*. *J Cell Biol* 127(5): 1173-1184.

Wilusz J, Shenk T, Takagaki Y, Manley JL (1990) A multicomponent complex is required for the AAUAAA-dependent cross-linking of a 64-kilodalton protein to polyadenylation substrates. *Mol Cell Biol* 10(3): 1244-1248.

Win TZ, Draper S, Read RL, Pearce J, Norbury CJ, Wang SW (2006) Requirement of fission yeast Cid14 in polyadenylation of rRNAs. *Mol Cell Biol* 26(5): 1710-1721.

Winey M, Culbertson MR (1988) Mutations affecting the tRNA-splicing endonuclease activity of *Saccharomyces cerevisiae*. *Genetics* 118(4): 609-617.

Winstall E, Sadowski M, Kuhn U, Wahle E, Sachs AB (2000) The *Saccharomyces cerevisiae* RNA-binding protein Rbp29 functions in cytoplasmic mRNA metabolism. *J Biol Chem* 275(29): 21817-21826.

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

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, Davis P, Feltwell T, Fraser A, Gentles S, Goble A, Hamlin N, Harris D, Hidalgo J, Hodgson G, Holroyd S, Hornsby T, Howarth S, Huckle EJ, Hunt S, Jagels K, James K, Jones L, Jones M, Leather S, McDonald S, McLean J, Mooney P, Moule S, Mungall K, Murphy L, Niblett D, Odell C, Oliver K, O'Neil S, Pearson D, Quail MA, Rabbinowitsch E, Rutherford K, Rutter S, Saunders D, Seeger K, Sharp S, Skelton J, Simmonds M, Squares R, Squares S, Stevens K, Taylor K, Taylor RG, Tivey A, Walsh S, Warren T, Whitehead S, Woodward J, Volckaert G, Aert R, Robben J, Grymonprez B, Weltjens I, Vanstreels E, Rieger M, Schafer M, Muller-Auer S, Gabel C, Fuchs M, Dusterhoft A, Fritze C, Holzer E, Moestl D, Hilbert H, Borzym K, Langer I, Beck A, Lehrach H, Reinhardt R, Pohl TM, Eger P, Zimmermann W, Wedler H, Wambutt R, Purnelle B, Goffeau A, Cadieu E, Dreano S, Gloux S, Lelaure V, Mottier S, Galibert F, Aves SJ, Xiang Z, Hunt C, Moore K, Hurst SM, Lucas M, Rochet M, Gaillardin C, Tallada VA, Garzon A, Thode G, Daga RR, Cruzado L, Jimenez J, Sanchez M, del Rey F, Benito J, Dominguez A, Revuelta JL, Moreno S, Armstrong J, Forsburg SL, Cerutti L, Lowe T, McCombie WR, Paulsen I, Potashkin J, Shpakovski GV, Ussery D, Barrell BG, Nurse P (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415(6874): 871-880.

Wood V, Harris MA, McDowall MD, Rutherford K, Vaughan BW, Staines DM, Aslett M, Lock A, Bahler J, Kersey PJ, Oliver SG (2012) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res* 40(Database issue): D695-699.

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

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

Xu F, Cohen SN (1995) RNA degradation in *Escherichia coli* regulated by 3' adenylation and 5' phosphorylation. *Nature* 374(6518): 180-183.

Xu F, Lin-Chao S, Cohen SN (1993) The *Escherichia coli* *pcnB* gene promotes adenylation of antisense RNAI of ColE1-type plasmids in vivo and degradation of RNAI decay intermediates. *Proc Natl Acad Sci U S A* 90(14): 6756-6760.

Yamanaka S, Yamashita A, Harigaya Y, Iwata R, Yamamoto M (2010) Importance of polyadenylation in the selective elimination of meiotic mRNAs in growing *S. pombe* cells. *EMBO J* 29(13): 2173-2181.

Yamashita A, Chang TC, Yamashita Y, Zhu W, Zhong Z, Chen CY, Shyu AB (2005) Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat Struct Mol Biol* 12(12): 1054-1063.

Yang PK, Hoareau C, Froment C, Monsarrat B, Henry Y, Chanfreau G (2005) Cotranscriptional recruitment of the pseudouridylsynthetase Cbf5p and of the RNA binding protein Naf1p during H/ACA snoRNP assembly. *Mol Cell Biol* 25(8): 3295-3304.

Yang PK, Rotondo G, Porras T, Legrain P, Chanfreau G (2002) The Shq1p.Naf1p complex is required for box H/ACA small nucleolar ribonucleoprotein particle biogenesis. *J Biol Chem* 277(47): 45235-45242.

Yang Q, Gilmartin GM, Doublet S (2010) Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc Natl Acad Sci U S A* 107(22): 10062-10067.

Yoon JH, Gorospe M (2016) Identification of mRNA-Interacting Factors by MS2-TRAP (MS2-Tagged RNA Affinity Purification). *Methods Mol Biol* 1421(15-22).

Yu Y, Ren JY, Zhang JM, Suo F, Fang XF, Wu F, Du LL (2013) A proteome-wide visual screen identifies fission yeast proteins localizing to DNA double-strand breaks. *DNA Repair (Amst)* 12(6): 433-443.

Yu YT, Meier UT (2014) RNA-guided isomerization of uridine to pseudouridine--pseudouridylation. *RNA Biol* 11(12): 1483-1494.

Yue Z, Maldonado E, Pillutla R, Cho H, Reinberg D, Shatkin AJ (1997) Mammalian capping enzyme complements mutant *Saccharomyces cerevisiae* lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. *Proc Natl Acad Sci U S A* 94(24): 12898-12903.

Yuryev A, Patturajan M, Litingtung Y, Joshi RV, Gentile C, Gebara M, Corden JL (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A* 93(14): 6975-6980.

Zamft B, Bintu L, Ishibashi T, Bustamante C (2012) Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc Natl Acad Sci U S A* 109(23): 8948-8953.

Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J (1999) Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol Cell Biol* 19(11): 7461-7472.

Zhang H, Rigo F, Martinson HG (2015) Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Mol Cell* 59(3): 437-448.

Zhang Z, Fu J, Gilmour DS (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev* 19(13): 1572-1580.

Zhang Z, Gilmour DS (2006) Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell* 21(1): 65-74.

Zhao R, Kakihara Y, Gribun A, Huen J, Yang G, Khanna M, Costanzo M, Brost RL, Boone C, Hughes TR, Yip CM, Houry WA (2008) Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. *J Cell Biol* 180(3): 563-578.

Zheng D, Tian B (2014) RNA-binding proteins in regulation of alternative cleavage and polyadenylation. *Adv Exp Med Biol* 825(97-127).

Zhou Y, Zhu J, Schermann G, Ohle C, Bendrin K, Sugioka-Sugiyama R, Sugiyama T, Fischer T (2015) The fission yeast MTREC complex targets CUTs and unspliced pre-mRNAs to the nuclear exosome. *Nat Commun* 6(7050).

Zuo Y, Deutscher MP (2001) Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res* 29(5): 1017-1026.