

Université de Montréal

**Caractérisation du domaine C-terminal de l'ARN polymérase II et  
de la phosphatase Glc7 dans la terminaison transcriptionnelle chez  
*Saccharomyces cerevisiae***

par

Pierre Collin

Programmes de Biologie moléculaire

Faculté de Médecine

Thèse présentée à la Faculté de Médecine  
en vue de l'obtention du grade de *Philosophiae doctor* (Ph.D.)  
en Biologie moléculaire

Décembre 2018

© Pierre Collin, 2018



## Résumé

La terminaison transcriptionnelle de l'ARN polymérase II (ARNPII) est l'étape de la transcription la moins bien comprise. Deux modèles ont toutefois été proposés pour la levure *Saccharomyces cerevisiae*. Pour les gènes codants, il est suggéré que la phosphorylation de la Ser2 du domaine C-terminal (CTD) de l'ARNPII, en concomitance avec la déphosphorylation de la Tyr1, permet le recrutement des facteurs de terminaison Rtt103 et Pcf11 (complexe CPF-CF). Pour les ARN non-codants, la terminaison nécessiterait la phosphorylation de la Ser5 du CTD pour le recrutement du facteur de terminaison Nrd1 (complexe Nrd1-Nab3-Sen1 (NNS)). Même si le rôle du complexe CPF-CF est souvent présenté comme étant exclusif aux gènes codants, il a aussi été démontré que certains de ses constituants sont impliqués dans la terminaison des ARN non-codants. Il a notamment été proposé qu'une des phosphatases du CPF-CF, Glc7, soit impliquée dans l'activation de l'hélicase Sen1. Ces différents modèles de la terminaison transcriptionnelle comportent, cependant, des limitations. En effet, les mécanismes impliquant le CTD sont basés principalement sur l'utilisation de mutants des kinases et des phosphatases du CTD, car les mutants du domaine sont pour la plupart létaux. Puis, pour ce qui est du rôle de Glc7 dans l'activation de Sen1, aucune évidence *in vivo* ne le démontre.

Le projet principal de mon doctorat consistait donc à réévaluer, par des approches génomiques, l'implication du CTD de l'ARNPII dans la terminaison à l'aide de différents systèmes d'expression ectopique rendant possible l'utilisation de mutants du CTD. Comme attendu, les résultats obtenus démontrent des défauts globaux de terminaison aux gènes codants avec les mutants de la Ser2 et de la Thr4. Étonnamment, aucun défaut de terminaison n'est observé aux ARN non-codants avec le mutant de la Ser5. Les défauts de terminaison à ces ARN sont plutôt observés avec le mutant de la Tyr1, tandis que seulement un nombre limité de gènes codants est affecté par ce mutant. De plus, nous avons découvert que la Tyr1 est essentielle pour l'établissement de pauses de l'ARNPII dans la région 5' des gènes et que ces événements sont essentiels à une terminaison efficace aux ARN non-codants à l'aide du NNS.

Le deuxième projet consistait, quant à lui, à déterminer le rôle de la déphosphorylation de Sen1 par la phosphatase Glc7. Les résultats obtenus démontrent que Sen1 est une enzyme stable, mais que cette stabilité dépend de Glc7. De plus, l'effet de Glc7 sur Sen1 est dépendant de l'activité catalytique de Glc7 et du protéasome. Enfin, ce rôle de Glc7 est indépendant de son interaction avec les complexes CPF-CF ou APT.

En somme, les résultats obtenus durant mes études permettent d'effectuer une mise à jour importante quant au rôle que joue le CTD dans la terminaison transcriptionnelle et celui que joue Glc7 dans la terminaison des ARN non-codants.

**Mots clés :** Terminaison transcriptionnelle, ARN polymérase II, CTD, Tyrosine 1, pause, Nrd1, Sen1, Glc7, Pcf11, Rtt103

## Abstract

Transcriptional termination of RNA polymerase II (RNAPII) is the least understood step of transcription. Two models have been proposed in *Saccharomyces cerevisiae*. At protein-coding genes, it is proposed that the phosphorylation of Ser2 of the RNAPII C-terminal domain (CTD), in concomitance with the dephosphorylation of Tyr1, triggers the recruitment of termination factors Rtt103 and Pcf11 (CPF-CF complex). At non-coding RNAs, it is proposed that Ser5 phosphorylation is required for Nrd1 recruitment (Nrd1-Nab3-Sen1 complex (NNS)). Although the CPF-CF complex is often presented as exclusive to protein-coding genes, some of its components have also been shown to be involved in the termination of non-coding RNAs. Notably, it has been proposed that one of the two phosphatases present in CPF-CF, Glc7, is involved in the activation of the helicase Sen1. These different models of termination have, however, limitations. Indeed, the mechanisms involving the CTD are mainly based on the use of mutants of CTD kinases and phosphatases since most CTD mutants are sick or lethal. In addition, the role of Glc7 in the activation of Sen1 has not been demonstrated *in vivo*.

The main project of my Ph.D. was therefore to reinvestigate, using genomic approaches, the role of the RNAPII CTD in termination using different ectopic expression systems making possible the use of CTD mutants. As expected, the results obtained demonstrate a global termination defect at protein-coding genes with the Ser2 and Thr4 mutants. Surprisingly, however, no termination defect was observed at non-coding RNAs with the Ser5 mutant. Defects in non-coding RNA terminations were rather observed with the Tyr1 mutant, whereas only a limited number of protein-coding genes was affected in this mutant. In addition, we found that Tyr1 is essential for RNAPII pausing in the 5' region of genes and that these events are essential to allow efficient NNS-dependent termination.

The second project consisted in determining the role of Sen1 dephosphorylation by the phosphatase Glc7. The results obtained demonstrate that Sen1 is a stable enzyme, but also that its stability requires Glc7. In addition, the effect of Glc7 on Sen1 is dependent on the catalytic

activity of Glc7 and is proteasome-dependent. Finally, this role of Glc7 is independent of its interaction with CPF-CF or APT.

In conclusion, the results obtained during my PhD allowed for an important update of the role of the CTD in the transcriptional termination and the mechanism through which Glc7 regulates the termination of non-coding RNAs.

**Keywords:** Transcription termination, RNA polymerase II, CTD, Tyrosine 1, pausing, Nrd1, Sen1, Glc7, Pcf11, Rtt103

## Table des matières

Résumé.....	i
Abstract.....	iii
Table des matières.....	v
Liste des figures.....	viii
Liste des tableaux.....	ix
Liste des abréviations .....	x
Remerciements .....	xii
<b>CHAPITRE 1: Introduction .....</b>	<b>1</b>
<b>1.1 Bref histoire de la génétique moléculaire .....</b>	<b>3</b>
<b>1.2 ARN polymérase II .....</b>	<b>6</b>
1.2.1 Structure .....	7
<b>1.3 Domaine C-terminal (CTD) de l'ARNPII .....</b>	<b>11</b>
1.3.1 Caractérisation génétique .....	14
1.3.2 Structure .....	16
1.3.3 Caractéristique physico-chimique .....	18
1.3.4 Modifications post-traductionnelles .....	19
1.3.4.1 Cycle de phosphorylation .....	19
1.3.4.1.1 Kinases et phosphates .....	23
1.3.4.2 Autres .....	27
1.3.5 Protéines liant les différentes formes phosphorylées .....	28
<b>1.4 Transcription : Initiation et élongation .....</b>	<b>32</b>
1.4.1 Initiation .....	32
1.4.2 Élongation .....	38
1.4.2.1 Pause au promoteur proximal .....	39
1.4.2.2 Facteurs d'élongation .....	42
1.4.2.2.1 DSIF .....	43
1.4.2.2.2 TFIIF .....	44
1.4.2.2.3 TFIIS .....	44
1.4.2.2.4 PAF1C .....	45
1.4.2.2.5 FACT .....	45
1.4.2.2.6 Spt6 .....	46
1.4.2.3 Évènements co-transcriptionnelles .....	47
1.4.2.3.1 Coiffe des ARN .....	47
1.4.2.3.2 Épissage .....	48
<b>1.5 Transcription: Terminaison .....</b>	<b>49</b>
1.5.1 Terminaison transcriptionnelle chez <i>S. cerevisiae</i> .....	50
1.5.2 Terminaison transcriptionnelle chez les mammifères .....	57
1.5.3 Évènements co-transcriptionnelles .....	61
1.5.3.1 Clivage et poly-adénylation des ARNm .....	61
1.5.3.2 Exportation des ARNm .....	61
<b>1.6 Problématiques, objectifs et méthodologies .....</b>	<b>62</b>

<b>CHAPITRE 2: La Tyr1 du CTD de l'ARN polymérase II est requise pour une terminaison transcriptionnelle optimale par le complexe Nrd1-Nab3-Sen1 .....</b>	<b>67</b>
<b>2.1 Résumé.....</b>	<b>68</b>
<b>2.2 Contribution.....</b>	<b>68</b>
<b>2.3 Manuscrit.....</b>	<b>69</b>
2.3.1 Abstract .....	70
2.3.2 Introduction .....	71
2.3.3 Results .....	73
2.3.3.1 A dual-tag system to explore the role of the RNAPII CTD in transcription... ..	73
2.3.3.2 Ser2 and Thr4 are globally required for efficient termination by the CPF-CF pathway .....	74
2.3.3.3 Tyr1 is required for the efficient termination of a limited number of protein-coding genes.....	76
2.3.3.4 Tyr1 mutations lead to massive readthrough at non-coding RNA genes .....	77
2.3.3.5 NNS termination factor recruitment does not require their interaction with the CTD .....	79
2.3.3.6 Tyr1 is required for RNAPII pausing in the 5' end of genes .....	79
2.3.3.7 Termination defects in the Tyr1 mutant are suppressed by slow RNAPII .....	81
2.3.4 Discussion .....	81
2.3.5 Acknowledgments .....	84
2.3.6 References .....	85
2.3.7 Figures .....	94
2.3.8 Figure legends .....	101
2.3.9 Star methods .....	105
2.3.9.1 Key resources table.....	105
2.3.9.2 Experimental model and subject details .....	107
2.3.9.3 Detailed methods .....	108
2.3.9.4 Quantification and statistical analysis .....	117
2.3.10 Supplemental Information .....	122
2.3.10.1 Figures .....	122
2.3.10.2 Figure legends.....	129
2.3.10.3 Tables.....	132
<b>CHAPITRE 3: La stabilité de l'hélicase Sen1 est régulée par la phosphatase Glc7 de façon dépendante du protéasome .....</b>	<b>143</b>
<b>3.1 Résumé.....</b>	<b>144</b>
<b>3.2 Contribution des auteurs .....</b>	<b>144</b>
<b>3.3 Manuscrit.....</b>	<b>145</b>
3.3.1 Abstract .....	145
3.3.2 Introduction .....	146
3.3.3 Results .....	148
3.3.3.1 Glc7 regulates the stability of Sen1 .....	148
3.3.3.2 Glc7 regulates Sen1's fate independently of its interaction with CPF-CF or APT .....	150
3.3.4 Discussion .....	150
3.3.5 Acknowledgments .....	152



3.3.6	Reference.....	152
3.3.7	Figures.....	157
3.3.8	Figure legends.....	159
3.3.9	Materials and Methods.....	160
3.3.10	Supplemented information.....	162
3.3.10.1	Tables.....	162
<b>CHAPITRE 4:</b>	<b>Discussion et conclusion.....</b>	<b>167</b>
4.1	Le rôle du CTD dans la terminaison des gènes codants.....	168
4.2	Le rôle du CTD dans la terminaison des gènes courts non-codants.....	170
4.3	Le rôle de la déphosphorylation de Sen1 par la phosphatase Glc7.....	176
4.4	Conclusion.....	178
<b>Bibliographie.....</b>		<b>179</b>

## Liste des figures

Figure 1.1 Structure de l'ARNPII. ....	8
Figure 1.2 Schématisation du modèle représentant le cycle de l'addition des nucléotides à l'ARN naissant par l'ARNPII et de sa translocation.....	11
Figure 1.3 Schématisation du domaine C-terminal (CTD) de l'ARNPII.....	12
Figure 1.4 Représentation du CTD de l'ARNPII chez <i>S. cerevisiae</i> et chez l'humain.....	14
Figure 1.5 Cycle de la phosphorylation du CTD de l'ARNPII.....	21
Figure 1.6 Schématisation de la formation du complexe de pré-initiation de la transcription. ....	36
Figure 1.7 Schématisation des mécanismes de terminaison transcriptionnelle chez <i>S. cerevisiae</i> . ....	51
Figure 1.8 Schématisation des mécanismes de terminaison transcriptionnelle chez les mammifères.....	59
Figure 2.1 A dual-tag system to explore the role of the RNAPII CTD in transcription.....	101
Figure 2.2 CTD Ser2 and Thr4 are ubiquitously required for efficient termination at protein-coding genes, while Tyr1 has a gene-specific role. ....	101
Figure 2.3 CTD Tyr1 is important for snoRNA termination.....	102
Figure 2.4 Tyr1 phosphomimetics disrupt snoRNA termination.....	102
Figure 2.5 Nrd1 and Sen1 are recruited to snoRNA loci independently of their interaction with the CTD.....	102
Figure 2.6 Pausing in the 5' end of genes is decreased in Y1F cells. ....	103
Figure 2.7 Impairing elongation rescues the termination defects of Y1F cells. ....	103
Figure S2.1 Related to Figure 2.1. ....	129
Figure S2.2 Related to Figure 2.2.....	129
Figure S2.3 Related to Figure 2.2.....	130
Figure S2.4 Related to Figure 2.3.....	131
Figure S2.5 Related to Figure 2.5.....	131
Figure S2.6 Related to Figure 2.6.....	131
Figure S2.7 Related to Figure 2.7.....	131
Figure 3.1 Glc7-regulated Sen1 degradation via the proteasome.....	159
Figure 3.2 The role of Glc7 in Sen1 stability is independent of its interaction with CPF-CF or APT.....	159

## Liste des tableaux

Table 1.1 Kinases et Phosphatases du CTD ayant des évidences <i>in vivo</i> .....	27
Table S2.1 Gene lists from NET-seq experiments.....	132
Table S2.2 Gene lists from ChIP-chip experiments.....	132
Table S2.3 Pause sites identified in this study.....	132
Table S2.4 Yeast strains used in this study.....	133
Table S2.5 Plasmids used in this study.....	137
Table S2.6 Sequence of each CTD mutant used in this study. ....	138
Table S2.7 Oligonucleotides used in this study.....	142
Table S3.1 Yeast strains used in this study.....	162
Table S3.2 Plasmids used in this study.....	164
Table S3.3 Oligonucleotides used in this study.....	165

## Liste des abréviations

6AU	: 6-azauracile
A	: Alanine
ADN	: Acide déoxyribonucléique
Ala	: Alanine
Arg	: Arginine
ARN	: Acide ribonucléique
ARNm	: ARN messenger
ARNnc	: ARN non-codant
ARNPI	: ARN Polymérase I
ARNPII	: ARN Polymérase II
ARNPIII	: ARN Polymérase III
Asn	: Asparagine
ATP	: Adénosine triphosphate
ChIP-chip	: Immunoprécipitation de la chromatine hybridée à des micropuces / <i>Chromatin immunoprecipitation on chip</i>
ChIP-seq	: Immunoprécipitation de la chromatine couplée aux séquençage à haut-débit / <i>Chromatin immunoprecipitation sequencing</i>
CID	: Domaine d'interaction au CTD / <i>C-terminal interaction domain</i>
CPF	: <i>Cleavage and polyadenylation factor</i>
cryo-EM	: Cryo-microscopie électronique
CTD	: Domaine en C-terminal / <i>C-terminal domain</i>
C-ter	: C-terminal
CUT	: <i>Cryptic unstable transcript</i>
D	: Asparagine
DC	: Dichroïsme circulaire
dNTP	: Désoxyribonucléotides
DNase	: Désoxyribonucléase
DSIF	: <i>DRB-sensitivity-inducing factor</i>
E	: Acide glutamique
Glu	: Acide glutamique
F	: Phénylalanine
GTF	: Facteur général de la transcription
HAT	: Histone acétyltransférase
IP	: Immunoprécipitation
K	: Lysine
Lys	: Lysine
NELF	: <i>Negative elongation factor</i>
NET-seq	: <i>Native elongating transcript sequencing</i>
NFR	: <i>Nucleosome free region</i>
NNS	: Nrd1-Nab3-Sen1
NTP	: Nucléotide triphosphate
N-ter	: N-terminal
ORF	: <i>Open reading frame</i>

P	: Proline
PCR	: Réaction en chaîne par polymérase
Phe	: Phénylalanine
PIC	: Complexe de pré-initiation
Poly(A)	: Poly-adénylation
Pro	: Proline
P-TEFb	: <i>Positive transcription elongation factor B</i>
qPCR	: Réaction en chaîne par polymérase quantitative
R	: Arginine
RMN	: Résonance magnétique nucléaire
RNase	: Ribonucléase
RT-qPCR	: <i>Quantitative reverse transcription polymerase chain reaction</i>
S	: Sérine
Ser	: Sérine
Ser-P	: Sérine phosphorylée
snoRNA	: Petit ARN nucléolaire
snRNA	: Petit ARN nucléaire
T	: Thréonine
TAF	: Facteur associé au TBP
TBP	: <i>TATA-binding protein</i>
Thr	: Thréonine
Thr-P	: Thréonine phosphorylée
TSS	: Site d'initiation de la transcription
TTS	: Site de terminaison de la transcription
Tyr	: Tyrosine
Tyr-P	: Tyrosine phosphorylée
UAS	: <i>Upstream activator site</i>
V	: Valine
Val	: Valine
Y	: Tyrosine

## Remerciements

Pour commencer, j'aimerais remercier les membres du jury pour l'évaluation de ma thèse de doctorat : Dr Jacques Drouin, Dr Daniel Zenklusen, Dr François Bachand et Dr Francis Rodier. J'aimerais aussi remercier Dr Jacques Côté pour avoir participé comme membre à mon comité de suivi.

Je remercie également Dr François Robert de m'avoir pris sous son aile et de m'avoir fait confiance durant ces années. Ce fût très agréable de vous côtoyer autant au sujet de la science que pour Star Wars! Merci de m'avoir donné l'opportunité de participer à différents congrès à l'étranger afin de présenter mes recherches.

J'aimerais aussi remercier les différents membres du laboratoire. Célia pour nos discussions autant scientifiques que politiques. Merci de m'avoir enseigné les différentes techniques du laboratoire et de m'avoir partagé ta passion pour la science. J'aimerais aussi remercier Anne et Nicole pour nos nombreuses discussions et pour leur écoute. Merci aussi à Fanny pour ta présence, ton écoute et ta bonne humeur, sans toi le temps aurait été long! Bonne chance pour ta future carrière et pour ton bébé. On va avoir l'occasion de se revoir! J'aimerais aussi remercier le groupe de Marlene Oeffinger, plus précisément Christian Trahan pour son aide et ses conseils techniques.

Je remercie aussi mes amis et ma famille pour leur soutien, surtout Étienne pour ses conseils et pour nos discussions scientifiques et historiques autour d'une bière. Bonne chance pour ta carrière de chercheur.

Enfin, j'aimerais remercier ma conjointe Lucie pour sa patience durant toutes ces années. Merci de m'avoir écouté, supporté et d'avoir toujours été là pour moi peu importe les moments. Tu as été ma première source de motivation et d'encouragement dans les situations difficiles. Merci aussi à mon petit Albert qui réussit toujours à me mettre de bonne humeur et qui a été, avec Lucie, le plus beau cadeau que la vie pouvait me donner!

# **CHAPITRE 1: Introduction**

Le bagage génétique, constitué d'acide désoxyribonucléique (ADN), correspond à l'information nécessaire pour la constitution d'une cellule, à son maintien et à son adaptation face à son environnement, et ce, peu importe l'organisme vivant. L'ADN comprend différentes informations permettant l'obtention ultime de protéines et d'acide ribonucléique (ARN) non-codants. Pour les protéines, un intermédiaire entre l'ADN et ces molécules doit être synthétisé, l'ARN messager (ARNm). Ainsi, l'ADN doit être transcrit en ARNm dans le noyau pour être ensuite traduit en protéine dans le cytoplasme. Toutes les étapes impliquées dans ces processus sont hautement contrôlées et conservées. La transcription génétique est, quant à elle, effectuée par une enzyme, l'ARN polymérase II (ARNPII) et elle est divisée en trois grandes étapes : l'initiation, l'élongation et la terminaison.

L'initiation correspond au recrutement de l'ARNPII au promoteur des gènes par différents facteurs de transcription. Une fois la machinerie mise en place, la transcription débute et entre ensuite dans la phase d'élongation. Lorsque le gène est enfin transcrit, la polymérase doit s'arrêter et terminer. Au fil des ans, beaucoup d'informations sur les séquences génétiques, les protéines impliquées et leurs structures ont été déterminées. De plus, le domaine en C-terminal (CTD) de l'ARNPII a aussi été démontré pour subir différentes modifications durant le cycle de la transcription afin de recruter les protéines impliquées dans chacune des étapes.

Contrairement à l'initiation et à l'élongation, la terminaison est l'étape la moins bien comprise. Chez la levure *Saccharomyces cerevisiae*, il existe deux mécanismes impliquant le CTD, un pour les gènes codants pour des protéines (conservé chez l'humain) et un pour les gènes courts non-codants. Chacun de ces mécanismes utilisent des protéines qui leur sont propres. Cependant, des évidences démontrent qu'il y a des protéines impliquées dans la terminaison des gènes codants qui le sont aussi pour les gènes courts non-codants, comme la phosphatase Glc7, sans que l'on connaisse leurs rôles.

Le sujet de cette thèse porte donc sur la caractérisation du CTD de l'ARNPII dans la terminaison transcriptionnelle chez la levure *S. cerevisiae*. Il est déjà bien établi que ce domaine joue un rôle important dans cette étape. Toutefois, aucune étude systémique utilisant



des mutants directs du CTD n'a été effectuée, car ces mutants sont létaux. Pour ce faire, nous avons développé deux systèmes *in vivo* d'expression ectopique permettant leurs utilisations et leurs études. De plus, il sera aussi question de la caractérisation de la phosphatase Glc7 dans la terminaison des gènes courts non-codants.

## 1.1 Bref histoire de la génétique moléculaire

Les débuts de la génétique remontent au 19<sup>e</sup> siècle avec le biologiste Charles Darwin qui proposa sa théorie de l'évolution basée sur la sélection naturelle. À partir de cet instant, les premières notions de l'hérédité furent élaborées [1]. Cette théorie révolutionnaire, très critiquée à l'époque, est basée sur l'observation d'animaux et de plantes qui a permis à Darwin d'émettre l'hypothèse qu'une même espèce située dans des environnements différents va évoluer pour s'y adapter à l'aide d'acquisitions. Ces dernières sont transmissibles lors de la reproduction, ce qui met de l'avant la notion de l'hérédité. Durant le même siècle, un autre personnage important de la théorie de l'hérédité fut le moine tchèque, Gregor Mendel. Celui-ci élaborera la base de la génétique moderne selon trois théories qui portent dorénavant son nom, la théorie mendélienne, fondée sur l'observation de traits physiologiques de petits pois. Il propose alors l'aspect récessif et dominant des allèles. Toutefois, ses différentes découvertes furent oubliées et seulement redécouvertes, indépendamment, par les trois botanistes Carl Correns, Erich von Tschermak-Seysenegg et Hugo de Vries en 1900 [2]. Ce dernier suggéra sa théorie des mutations pour expliquer celle de l'évolution et proposa que les mutations peuvent être héritées par la descendance selon les lois de Mendel. En 1902, Theodor Boveri et Walter Sutton émirent leurs théories chromosomiques de l'hérédité basées sur l'expérimentation. Lors de la division cellulaire, ils firent l'observation que les chromosomes étaient divisés de façon égale entre les cellules filles et mères, et que les chromosomes de l'œuf fécondé sont une combinaison de ceux de la mère et du père [3]. Ces observations furent ensuite confirmées par Thomas Morgan. À l'aide de la mouche à fruit *Drosophila melanogaster*, Morgan mit en relation le gène avec le chromosome basé sur l'observation de la couleur des yeux des mouches femelles et mâles, et de leur progéniture [4]. De plus, il fut le premier à observer que le gène responsable de la couleur des yeux se situait sur le chromosome sexuel des mouches. Il

est aussi important de noter qu'il fut le premier à utiliser la drosophile comme organisme modèle. Ses travaux importants sur la génétique lui permirent de décrocher le prix Nobel en physiologie et médecine en 1933.

Toutefois, ce n'est qu'à la fin de la Seconde guerre mondiale que les véritables bouleversements dans la compréhension de la génétique vont faire leur apparition grâce aux chimistes et aux physiciens. Dès lors, les scientifiques croyaient que le bagage génétique était constitué de protéines, du au fait que ces molécules sont très abondantes dans la cellule. Cependant, un dénommé Friedrich Miescher avait déjà fait une découverte importante au 19<sup>e</sup> siècle. En isolant le noyau de leucocytes, il avait déterminé que le principal constituant de cette organelle était acide et il lui donna le nom de nucléine, qui n'est en fait que le premier nom donné à l'ADN [5]. Une expérience clé dans la compréhension de l'ADN fut effectuée, en 1928, par Frederick Griffith, qui utilisa la salive de patients atteints de pneumonie afin d'y faire la culture de pneumocoques. Il constata que les bactéries virulentes étaient lisses et qu'après plusieurs repiquages, elles perdaient de leur virulence et devenait rugueuses. Il décida alors de tuer les cellules virulentes par chauffage et de les mélanger avec les bactéries non virulentes [6]. Cela a eu pour effet de transformer les cellules non virulentes en cellules virulentes. Cette expérience, appelée transformation, fut la première utilisation de cette technique de base couramment utilisée aujourd'hui dans les laboratoires de biologie moléculaire. Par la suite, les chimistes Avery, McLeod et McCarthy décidèrent de déterminer la nature de cette matière transformante. Ils la traitèrent avec différentes protéases pour en venir à la conclusion que ce n'est pas les protéines qui donnent ce phénotype, mais plutôt l'ADN, car lorsque cette matière est traitée avec de la DNase, il n'est plus possible de transformer les cellules [7]. Le tout a ensuite été confirmé, en 1952, par Alfred Hershey et Marta Chase qui, en utilisant des phages, ont déterminé que c'était seulement l'ADN de ces derniers qui pouvait permettre l'infection des bactéries et non leurs capsules protéiques [8]. Donc, l'ADN est la constituante des chromosomes responsables de l'hérédité.

De nombreuses choses restaient encore à découvrir à propos de l'ADN, comme ses composantes et sa structure. Une course était alors lancée pour la détermination de la structure. Linus Pauling (récipiendaire du prix Nobel en chimie en 1954) proposa une structure en hélice

à trois brins. Dans cette structure, les groupements phosphates forment le cœur de la structure tandis que les acides nucléiques pointent vers l'extérieur. Toutefois, ce qui rend impossible cette structure est la présence des charges négatives des groupements phosphates qui rend la structure très instable. Ce modèle a été cependant le point tournant pour les physiciens James Watson et Francis Crick. À l'aide des résultats de diffraction par rayons X de l'ADN obtenus par Rosalind Franklin, il était devenu évident pour Watson et Crick que la structure de l'ADN était une double hélice ayant les groupements phosphates vers l'extérieur et les bases vers l'intérieur. En 1949, Erwin Chargaff avait déjà démontré que quatre nucléotides constituaient l'ADN avec une quantité égale d'adénine et de thymine, ainsi que de cytosine et de guanine [9]. En se basant sur cette évidence, ils proposèrent que l'ADN fût constitué de paires de nucléotide, c'est-à-dire que l'adénine est complémentaire à la thymine et la cytosine avec la guanine. Ces travaux ont été publiés dans Nature en 1953 et cela leur a valu le prix Nobel de physiologie et de médecine en 1962 [10].

Déjà, à l'époque, il était connu que l'ADN était dans le noyau et que la synthèse des protéines avait lieu dans le cytoplasme. Cependant, l'intermédiaire n'était pas connu. Ce n'est qu'en 1961 que deux laboratoires démontrèrent l'existence d'un nouveau type d'ARN, soit l'ARNm, et l'interaction de cet ARN avec le ribosome [11, 12]. Ces résultats ont permis la naissance de la notion de transcription de l'ADN en ARNm et de la traduction de cet ARN en protéine. Par la suite, en 1956, Severo Ochoa et Arthur Kornberg ont fait l'isolement d'enzymes, appelés polymérases, capables de synthétiser l'ADN ou l'ARN [13, 14]. Ces découvertes leur ont valu l'obtention du prix Nobel en physiologie et en médecine en 1959.

Enfin, en 1969, Roeder et Rutter ont démontré que les cellules eucaryotes possédaient trois enzymes différentes capables de synthétiser l'ARN à partir de l'ADN, soit des ARN polymérase (ARNP), qui, contrairement aux procaryotes, n'en possèdent qu'un seul [15]. La nomenclature de ces polymérases fut établie selon leur ordre de fractionnement, soit I, II et III. Cette importante découverte fut par la suite confirmée en 1970 par le laboratoire de Pierre Chambon qui les nomma ARN polymérase A, B et C [16]. Des études subjacentes ont permis de démontrer que l'ARNP de type I est l'enzyme responsable de la transcription de l'ADN ribosomal; le type II est responsable de la transcription des gènes codants pour des protéines,

mais aussi de certains gènes non codants et finalement le type III permet de transcrire les ARN de transfert et l'ARN ribosomal 5S.

## 1.2 ARN polymérase II

Lorsque les différentes ARN polymérases avaient été découvertes par Roeder et Rutter en 1969 à l'aide de chromatographie DEAE-Sephadex, trois complexes avaient été élués selon leur poids moléculaire [15]. Le pourcentage de transcrit d'ARN provenant de ces polymérases varie grandement d'une enzyme à l'autre. L'ARN polymérase I (ARNPI) transcrit pour environ 75 % de l'ARN total de la cellule, 10 % pour l'ARNPII et 15 % pour l'ARN polymérase III [17]. Les trois polymérases présentent une grande homologie entre elles. Elles contiennent respectivement 14, 12 et 17 sous-unités. Les sous-unités Rpb5, 6, 8, 10 et 12 de l'ARNPII sont partagées avec l'ARNPI et III, tandis que les sept autres ont des formes paralogues chez les ARNPI et III [18]. Les sous-unités communes sont très bien conservées dans l'évolution. On en retrouve des paralogues chez les Archaea et cinq, dont Rpb1, se retrouvent chez la bactérie [18]. Toutefois, même si l'ARNPII ne transcrit que pour 10 % de l'ARN total, c'est l'enzyme la plus étudiée des trois.

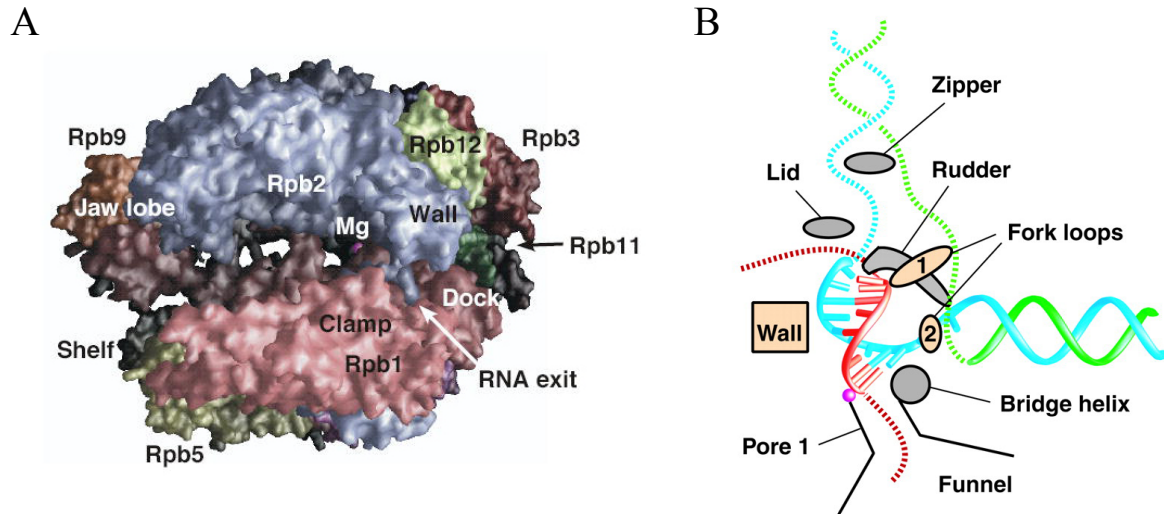
Dans le complexe protéique que forme l'ARNPII, les sous-unités Rpb4 et Rpb7 forment un dimère qui est en interaction transitoire avec le reste de l'enzyme [19]. Il a été démontré que ces deux sous-unités étaient nécessaires pour la réponse aux stress et pour l'initiation de la transcription, mais qu'elles ne l'étaient pas pour l'élongation [20, 21]. Environ 20 % des ARNPII contiennent Rpb4/Rpb7, ce qui démontre que le complexe est en association sous-stœchiométrique, et ce, même si Rpb4 est exprimé en plus grande quantité que les autres sous-unités [22-24]. Ce qui est intéressant avec cet hétérodimère, c'est qu'il est impliqué dans plusieurs événements autres que la transcription, et ce, même s'il n'est associé que transitoirement avec l'ARNPII. En effet, durant les dernières années, il a été démontré que le complexe était aussi impliqué dans l'exportation des ARNm, la stabilité des ARNm, la traduction et la réparation de l'ADN [25-28].

### 1.2.1 Structure

Au début des années 2000, un grand tour de force du laboratoire de Roger Kornberg a été effectué avec la détermination de la structure de l'ARNPII. Dans un premier temps, son équipe avait effectué une première cristallographie de l'enzyme, mais le cristal obtenu n'avait qu'une résolution de 5 Å, ce qui était beaucoup trop faible et, avec les limitations techniques, il n'était pas possible d'obtenir de meilleur résultat [29]. Toutefois, avec l'arrivée de Patrick Cramer comme stagiaire post-doctoral dans le laboratoire, une nouvelle approche technique a été effectuée et cela a permis d'effectuer une cristallographie de 3,5 Å [30]. Il est important de noter que ces premières structures comprennent seulement 10 sous-unités et que Rpb4 et Rpb7 sont manquants en raison du fait qu'ils sont en association sous-stœchiométrique avec le reste de l'enzyme. Fait intéressant, ces structures ont nécessité 10 000 L de culture de levure pour obtenir ces premiers cristaux. Par la suite, une série d'articles comprenant, par exemple, de meilleures résolutions, les douze sous-unités de l'ARNPII, l'enzyme en complexe avec l'ADN ou avec des facteurs de transcription ont été effectuées. Ces résultats ont grandement aidé et permis une meilleure compréhension de la transcription à l'échelle protéique et atomique.

La détermination de la structure de l'ARNPII a donc permis d'observer que la morphologie de l'enzyme est très similaire à celle observée avec l'ARN polymérase de la bactérie *Thermus aquaticus*, c'est-à-dire une forme en pince de crabe (Figure 1.1A) [31]. La surface de l'enzyme possède une charge totale négative, tandis que l'intérieur de la pince, correspondant à la surface de contact avec l'ADN, est de charge positive. L'enzyme peut être divisée en quatre modules mobiles [32]. Le premier module, appelé le *core*, représente la moitié de la masse de l'enzyme. Cette structure est composée de Rpb1 et de Rpb2 qui sont attachés l'un à l'autre par les sous-unités Rpb3, 10, 11 et 12 [30, 32]. Ce module permet de former un sillon d'environ 25 Å de diamètre, appelé le *cleft*, qui correspond à l'endroit où l'ADN s'insère et au site catalytique. Il est traversé par une structure de Rpb1 en forme d'hélice, appelée le *bridge*, et une extension dans le sens opposé du côté C-terminal de Rpb2, appelé le *wall*, qui empêche une extension linéaire de l'ADN dans le site actif. De plus, ces structures permettent l'apparition d'une région en forme d'entonnoir, appelé *funnel*, où un pore est visible sous le site actif. Ce dernier serre d'entrer aux nucléotides dans le site actif ainsi

que de porte de sortie pour l'ARN synthétisé lorsque l'ARNPII effectue une marche en arrière (*backtrack*).



**Figure 1.1 Structure de l'ARNPII.**

A) Représentation structurale de 10 sous-unités de l'ARNPII (excluant Rpb4 et Rpb7) et de ces quatre modules : *core*, *jaw lobe*, *clamp* et *shelf*. B) Représentation schématique de la structure de la bulle de transcription de l'ARNPII déterminée par cristallographie. L'ADN transcrit en ARN est représenté par la bande pleine de couleur bleue et son brin complémentaire en vert. L'ARN est, quant à lui, représenté en rouge. Les bandes pointillées représentent le positionnement possible des acides nucléiques qui n'ont pas été déterminés par cristallographie, contrairement aux bandes pleines. Les différents éléments de couleurs grises ou beiges représentent respectivement des boucles ou des extensions protéiques des sous-unités Rpb1 et Rpb2. Figure adaptée de Hahn, 2004 [20] et de Gnatt *et al.*, 2001 [33].

Le second module est le *clamp*, qui correspond au pouce de la pince de crabe. Cette structure, déjà visible avec une structure de faible résolution obtenue auparavant, est composée de Rpb1 et du C-terminal de Rpb2 [29]. Il est attaché au *cleft* via des régions *switch* de Rpb2, et de Rpb6, connu pour être phosphorylée, ce qui permet un mouvement de 14 Å du *clamp* afin que celui-ci se referme sur le *cleft* [34, 35]. Le *clamp* contient trois extensions en forme de boucle qui se situent en amont du site catalytique : le *rudder*, le *lid* et le *zipper*. De plus, à la base du *clamp*, on retrouve une cavité correspondant au canal de sortie de l'ARN. Enfin, la

détermination de la structure de Rpb4/7 a permis de démontrer que le dimère se lie à la base, en dessous du *clamp* et que Rpb7 permet de le barrer en position fermée, ce qui confirme le rôle du dimère dans l'initiation [21, 36, 37]. La région au bout de la pince appelée *jaws* est divisée en deux modules, soit le *jaw lobe* et le *shelf*. Ils comprennent une région de Rpb1, de Rpb9 et de Rpb2, et de Rpb5, de Rpb6 et de Rpb1, respectivement. Ces deux modules servent à agripper l'ADN lorsque le *clamp* est refermé sur celui-ci.

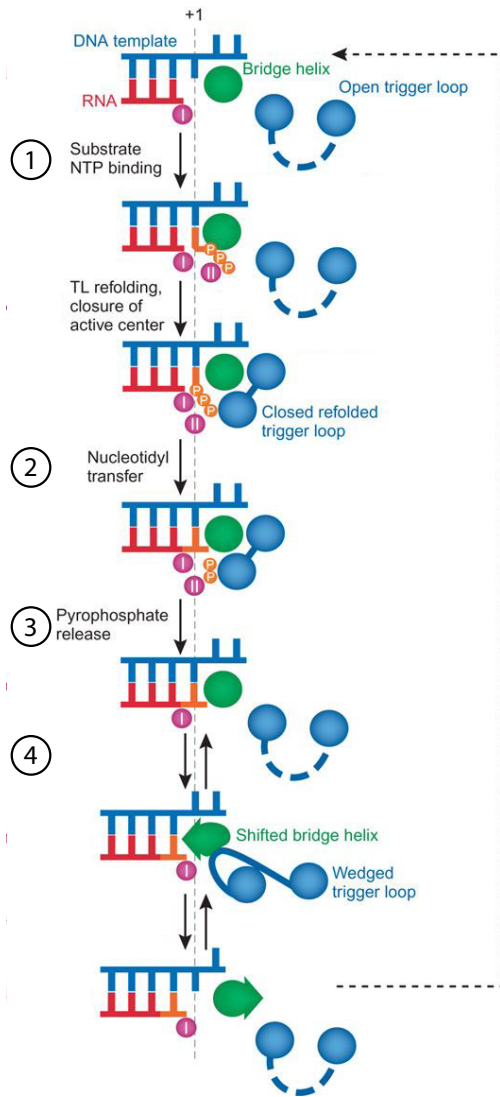
Des éléments intéressants du site actif ont aussi été observés et permettent de comprendre le mécanisme de transcription de l'ARNPII [33]. Le site actif se situe au fond de la pince via le *cleft* et implique plusieurs boucles avec des rôles très précis dans la formation et le maintien de la bulle de transcription. Cette dernière correspond à une longueur d'environ 15 paires de base (bp) d'ADN et de huit résidus pour l'hybride ADN-ARN. Comme on peut le voir dans la Figure 1.1B, cet hybride est en contact avec l'hélice *bridge* de Rpb1, le *rudder* du *clamp* et les boucles *fork* 1 et 2 de Rpb2. Le *bridge* permet le mouvement de translocation de l'ARNPII et le *rudder* prévient l'hybridation d'ARN et d'ADN, et pousse l'ARN vers la sortie qui se situe derrière le *clamp*. Les *fork* permettent de délimiter la région hybride. De plus, le *lid* agit comme barrière pour maintenir la séparation de l'ARN et de l'ADN et le *zipper* permet de maintenir la séparation des deux brins d'ADN. Puis lorsque l'ADN se situe en amont du *zipper*, il y a un rapprochement des deux brins. Il est aussi important de noter que toutes les interactions protéiques avec l'ADN ou l'ARN sont avec leurs groupements phosphates, ce qui ne permet aucune spécificité pour une base donnée. Comme attendu, deux ions  $Mg^{2+}$  sont aussi présents, soit le métal A et B [38]. Le métal A est en interaction stable avec trois résidus de Rpb1, c'est-à-dire D481, D483 et D485, et se situe dans la partie supérieur du pore [32]. Pour ce qui est du métal B, il est en interaction transitoire avec les résidus D481 et D483 de Rpb1, et avec E836 et D837 de Rpb2 et se situe dans la partie inférieure du pore [32, 39]. Ces résidus sont importants dans la réaction catalytique de l'ARNPII, car ils servent à stabiliser l'état transitoire durant la formation du lien phosphodiester [40].

La synthèse de l'ARN peut se décrire en quatre étapes en forme de cycle qui requièrent un réarrangement global du *bridge* et de la boucle *trigger* (Figure 1.2) [41]. La première consiste à la sélection du nucléotide triphosphate (NTP). La molécule s'accroche dans un site

transitoire qui est libéré, puis le *trigger* se déplace sur ce site afin de le fermer, ce qui a pour but de positionner le NTP [42-44]. Dans le cas d'un mauvais NTP, le *trigger* reste ouvert. Il est aussi important de noter que le site actif possède une affinité au moins 5000 fois plus grande pour les NTP que pour les désoxyribonucléotides (dNTP) [45]. La deuxième étape correspond à l'ajout du NTP à la fin 3' de l'ARN naissant à l'aide d'une réaction nucléophile par l'ion  $Mg^{2+}$  (métal A) [40, 46]. L'étape suivante se résume au relâchement d'un pyrophosphate du NTP à l'aide du métal B, ce qui a pour effet de forcer un réarrangement du *trigger* pour le mettre dans une position ouverte [47, 48]. Enfin, la dernière étape est la translocation de l'ARNPII sur l'ADN afin de libérer le bout 3' de l'ARN. Comme le site catalytique est bloqué par le nouveau NTP ajouté, il y a un changement de conformation à la fois du *bridge* et du *trigger* qui permet le mouvement [45, 49-52].

Dans les années qui ont suivi la découverte de la structure du site catalytique de l'ARNPII, différents mutants ont été créés et ont été démontrés pour avoir des effets sur la vitesse d'élongation. Trois mutations de la boucle *trigger* et une mutation dans la région de liaison entre le métal A et Rpb1 sont souvent utilisées pour ralentir ou augmenter la vitesse. La mutation H1085Y occasionne une perte d'un pont hydrogène avec le groupement phosphate du NTP dans le site actif [45, 49, 52]. Ce lien est nécessaire à la formation du lien phosphodiester lors de la catalyse. Donc, une mutation de ce résidu ralentit considérablement la vitesse de l'ARNPII [45, 53-55]. De plus, ce site est connu pour être l'endroit de liaison de la toxine  $\alpha$ -amanitine [45, 49]. Il a été démontré que deux autres résidus, G1097 et E1103, lorsque mutés, ont pour effet d'augmenter la vitesse d'élongation. La première mutation, G1097D, affecte l'hydrophobicité à l'entrée du site catalytique et la deuxième mutation, E1103G, affecte la fidélité de l'enzyme, car l'équilibre entre la position ouverte et fermée du site catalytique est débalancée pour la position ouverte [32, 54, 56-58]. Enfin, une mutation dans la région de liaison de Rpb1 avec le métal A, N488, a pour effet de ralentir la polymérase [32, 33, 58, 59].





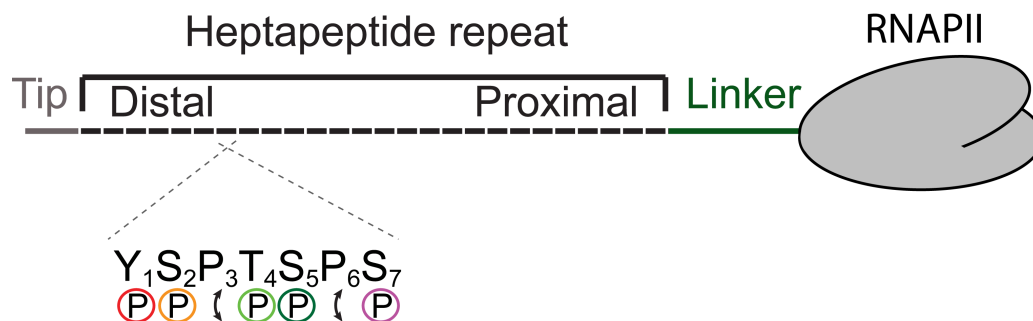
**Figure 1.2 Schématisation du modèle représentant le cycle de l'addition des nucléotides à l'ARN naissant par l'ARNPII et de sa translocation.**

Représentation des quatre étapes du processus : 1) la sélection du NTP, 2) l'ajout du NTP, 3) le relâchement d'un pyrophosphate et 4) la translocation de l'ARNPII. L'ADN est représenté en bleu, l'ARN, en rouge et le NTP, en orange. L'hélice *bridge* et la boucle *trigger* de Rpb1 sont représentés en vert et en bleu, respectivement. Figure adaptée de Svetlov et Nudler, 2008 [41].

### 1.3 Domaine C-terminal (CTD) de l'ARNPII

L'ARNPII comporte un domaine unique, hautement conservé et essentiel à la viabilité cellulaire, appelé le *C-terminal domain* (CTD). Ce domaine, qui a l'apparence d'une queue, a été découvert dans les années 1980, soit pratiquement 20 ans après la découverte de l'enzyme [60, 61]. Le CTD est unique à l'ARNPII, on ne la retrouve donc pas chez les autres ARN polymérase. Comme il existe une très grande homologie entre les trois polymérase au niveau des sous-unités et du site catalytique, cela indique que ce domaine est apparu après la divergence et la spécificité de chacune des ARNP [62]. Il est présent chez tous les organismes

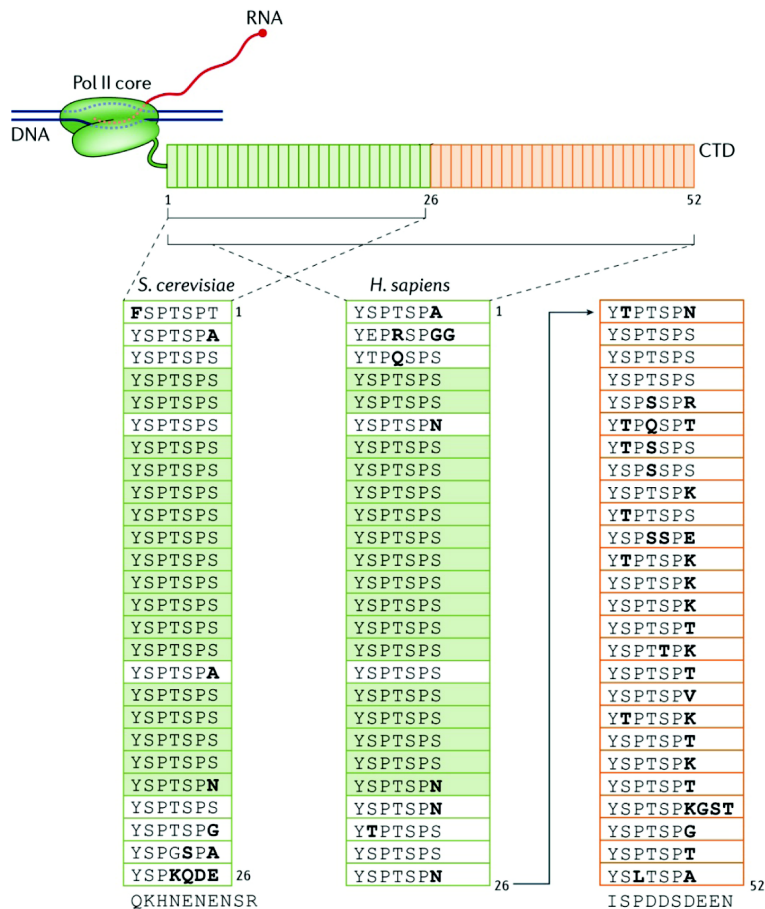
eucaryotes, mais absents des procaryotes et des Archaea [63]. Il est constitué de trois régions différentes, soit le *linker*, des répétitions d’heptapeptide et l’extrémité (*tip*) (Figure 1.3) [62, 64]. Le *linker*, qui est moins bien conservé que la répétition d’heptapeptide, est composé d’environ 80 acides aminés chez *S. cerevisiae* et il contient des motifs YSPx et SPxYx [62]. La partie centrale est, quant à elle, composée d’une série de répétitions d’un heptapeptide ayant comme séquence consensus (Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub>), où la tyrosine a été choisie comme étant la position un. Les différents résidus peuvent subir des modifications post-traductionnelles [65]. En effet, la Tyr1, Ser2, Thr4 Ser5 et Ser7 peuvent être phosphorylés ou N-glycosylés (sauf Tyr1), tandis que les deux prolines peuvent être isomérisées en position *cis* ou *trans* (Figure 1.3). Cette région de répétitions se divise en deux parties égales : la partie proximale et la partie distale. Les résidus Ser2Pro3 et Ser5Pro6 du CTD sont ceux qui sont les plus conservés, suivis de très près par la Tyr1, tandis que la Thr4 et la Ser7 le sont le moins [63]. Enfin, l’extrémité C-terminal du CTD peut être composée d’une séquence non consensus qui est impliquée dans la stabilité du domaine [66, 67].



**Figure 1.3 Schématisation du domaine C-terminal (CTD) de l'ARNPII.**

Le CTD est représenté par ses trois régions : le *linker*, la répétition d’heptapeptide et l’extrémité (*tip*). Chaque heptapeptide est composé d’une séquence consensus (Y<sub>1</sub>S<sub>2</sub>P<sub>3</sub>T<sub>4</sub>S<sub>5</sub>P<sub>6</sub>S<sub>7</sub>) qui varie en nombre et en dégénérescence selon l’organisme. La phosphorylation (P) est la modification post-traductionnelle la plus courante et la plus conservée pour les résidus Y<sub>1</sub>, S<sub>2</sub>, T<sub>4</sub>, S<sub>5</sub> et S<sub>7</sub>. Il en est aussi de même pour l’isomérisation en *cis* ou en *trans* des acides aminés P<sub>3</sub> et P<sub>6</sub>.

Ce qui différencie le CTD chez chacun des organismes est principalement la longueur du domaine et la dégénérescence de la séquence. Par exemple, la levure *S. cerevisiae* possède un CTD de 26 répétitions dont 19 contiennent la séquence consensus, tandis que celui des mammifères contient 52 répétitions où seulement la première moitié possède cette séquence consensus (Figure 1.4). Le reste est plutôt dégénéré, surtout à la position 7, et il a été proposé qu'il ait des tâches spécifiques [62]. D'autres organismes, comme la drosophile, ne contiennent que deux séquences consensus dans les 42 répétitions. Elle peut toutefois survivre lorsqu'on change son CTD pour celui de l'humain, ce qui suggère que l'aspect répétitif du CTD est aussi important, voire plus, que la séquence en elle-même [68]. Étonnamment, la séquence du CTD de l'humain, du rat et de la vache est identique. On retrouve aussi le même CTD chez la souris, mais avec une seule différence [64]. Celui du poisson-zèbre ne possède que seulement neuf différences en acide aminé par rapport à celui de l'humain. Donc lorsque l'on compare les différents CTD, on peut observer une corrélation entre la longueur du domaine, la présence de séquences non-consensus et les organismes les plus évolués [69]. Toutefois des exceptions sont aussi présentes, par exemple, le CTD de l'*Hydra magnipapillata* qui contient plus de 50 répétitions [63].



**Figure 1.4 Représentation du CTD de l'ARNPII chez *S. cerevisiae* et chez l'humain.**

Chaque répétition du CTD de la levure et de l'humain est représentée dans les rectangles. La levure contient 26 répétitions, tandis que l'humain en contient 52, où la première moitié est très similaire à celle de la levure. Les séquences identiques chez les deux organismes sont identifiées en vert. Les résidus qui diffèrent de la séquence consensus sont en caractères gras. Figure adaptée de Harlen *et al.*, 2017 [70].

Comme chaque répétition du CTD est encodée par 21 nucléotides, l'augmentation des répétitions du CTD peut être engendrée par une recombinaison des séquences répétées [71]. Si l'on clive le CTD de *S. cerevisiae* à 8 répétitions et que l'on laisse croître les cellules suffisamment, les cellules filles auront une augmentation du nombre de répétitions [72]. En somme, même si la séquence de beaucoup de CTD de différents organismes a été déterminée, celui de *S. cerevisiae*, qui possède un CTD ayant 19 séquences consensus sur 26 répétitions, est le candidat le plus étudié avec celui des mammifères.

### 1.3.1 Caractérisation génétique

Différentes études ont été effectuées dans les années qui ont suivi la découverte du CTD afin de déterminer la longueur minimale, l'espacement maximal entre les répétitions, les résidus essentiels et la différence entre la partie distale et proximale. Les premières études ont

démontré que l'ARNPII sans CTD est létale que ce soit chez la levure, les mammifères ou la drosophile [73-76]. Toutefois, une ARNPII sans ce domaine continue à posséder son activité catalytique, ce qui suggère que le CTD serait surtout nécessaire pour les processus co-transcriptionnels [77]. Tous ces organismes peuvent survivre avec des CTD tronqués [73-75]. Chez *S. cerevisiae*, la longueur minimale est de 8 répétitions, 16 des 29 répétitions sont nécessaires pour *S. pombe*, et 30 sur 52 pour les mammifères [72, 76, 78]. Pour ces derniers, un CTD possédant seulement les séquences non-consensus est létale [67, 76]. Donc, de façon générale, la moitié du CTD n'est pas nécessaire à la survie cellulaire.

Pour ce qui est des résidus essentiels, certaines mutations sont létales chez *S. cerevisiae*, comme lorsque tous les résidus Tyr1 sont remplacés par des phénylalanines (Y1F) [72]. Il en est aussi de même avec la mutation S2A et S5A [72]. Il est toutefois important de noter que l'étude de ces mutations effectuées par le laboratoire de Jeff Corden utilisait une expression ectopique à l'aide de plasmides. La même mutation du CTD S2A, lorsque intégrée au locus endogène, n'est pas létale. Un défaut de croissance sévère est toutefois observé [79]. Les mutations qui confèrent une imitation de la phosphorylation (phosphomimétique), comme S2E, S5E et S7E sont aussi létales [72, 80]. Toutefois, la mutation S7A et T4A permettent la viabilité cellulaire [81]. Étonnamment pour ce qui est de *S. pombe*, les mutations Y1F, S2A et S7E ne sont pas létales contrairement aux mutations S2E, S5A et S5E [79, 82-84]. Ces résultats proposent donc des rôles différents entre les deux levures pour les résidus Tyr1, Ser2 et Ser7. Chez l'humain les mutations phosphomimétiques de la Ser2, Ser5 et Ser7 sont aussi létales [64]. Il en est de même avec les mutations touchant la Thr4, c'est-à-dire T4A et T4V [85, 86]. La mutation Y1F est particulière chez l'humain. En effet, lorsque la mutation a lieu dans toutes les répétitions ayant une tyrosine, la stabilité du CTD est affectée et il y a une dégradation [87]. Afin de permettre une stabilité du domaine avec cette mutation, Shah *et al.* ont créé quatre différents mutants de la Tyr1, où le CTD a été divisé en quatre parties qui contiennent des mutations Y1F sur différentes sections: YFFF, YYFF, FYYF et YFFY [88]. De ces quatre mutants, seulement YYFF et FYYF sont viables. Cela suggère donc que les parties proximales et distales jouent un rôle différent dans le processus de la transcription. Enfin, toutes les mutations en alanine ou en glycine touchant les prolines en position trois ne supportent pas la viabilité cellulaire [64]. Étonnamment, la même

mutation en glycine pour la proline en position six est viable chez les trois organismes, ce qui démontre que les deux prolines ont des rôles complètement différents.

Comme il est indiqué dans la Figure 1.3, la région répétitive du CTD peut être divisée en deux parties selon la distance avec le *linker*. Une question se pose alors : est-ce que la région proximale joue le même rôle que celle distale? Afin de répondre à cette question, West *et al.* ont créé différents mutants [72]. Dans un premier temps, ils ont muté chaque Ser2 du CTD de *S. cerevisiae* en alanine seulement dans la région proximale, ce qui n'est pas létale pour la cellule. Étonnamment, lorsque les mutations sont effectuées dans la partie distale, il y a mort cellulaire. De plus, le même exercice a été effectué avec la mutation S5E et les mutations dans la partie distale sont viables, tandis que celles dans la partie proximale sont létales. Ces résultats démontrent donc que les différentes régions répétitives du CTD ont des rôles différents.

D'autres études ont aussi caractérisé l'espacement entre les différentes séquences consensus chez *S. cerevisiae* [72, 89, 90]. Différentes mutations avaient été effectuées afin d'ajouter un nombre varié d'alanines entre les différentes répétitions. Ces expériences ont permis de démontrer que peu importe le nombre d'alanines, ceci occasionne la mort cellulaire. En fait, pour permettre la viabilité cellulaire, ils ont déterminé qu'une unité fonctionnelle du CTD est nécessaire : YSPTSPS-YSPT. Puis, peu importe le nombre d'alanines entre ces unités, c'est-à-dire entre 1 et 5, la viabilité est supportée. De plus, des études chez *S. pompe* ont démontré qu'un CTD avec sept répétitions de la séquence YSPTSPS-YSPTAAA est viable, suggérant ainsi que ce ne sont pas tous les Ser5Pro6 de chaque répétition qui sont nécessaires [83].

### 1.3.2 Structure

La cristallisation de l'ARNPII a pu être effectuée pour les douze sous-unités de l'enzyme. Toutefois, le domaine CTD de l'ARNPII n'a pu être déterminé à cause de sa grande flexibilité et de son caractère désordonné [32]. Seulement la base du *linker* a pu être détectée.

Celui-ci se situe tout près du pore correspondant au canal de sortie de l'ARN. Il forme une hélice et possède une interaction avec la sous-unité Rpb7. Pour ce qui est du reste du domaine, rien n'est détectable. On sait toutefois que le CTD possède deux motifs SPXX, connus pour permettre la formation de structure en coude  $\beta$  ( $\beta$ -turn) [91]. Afin de déterminer une éventuelle structure du CTD, des expériences de résonance magnétique nucléaire (RMN), de dichroïsme circulaire (DC) et de microscopie électronique ont été effectuées. Dans un premier temps, des expériences de RMN avec une répétition du CTD en suspension dans l'eau ont permis d'observer un mélange de structures en coude  $\beta$  et de conformations non structurées [92]. D'autres expériences en RMN et en DC avec, cette fois-ci, un peptide représentant huit répétitions du CTD, ont permis d'observer une faible quantité de structure en coude  $\beta$  dans l'eau [93-95]. De plus, la présence d'hélice polyproline est aussi détectable à la même proportion que les structures en coude  $\beta$  [95]. Enfin, des analyses en microscopie électronique permettent d'observer une région à faible densité qui est attribuée à un CTD désordonné de 100 Å [96].

Basé sur ces observations, quatre structures du CTD ont été proposées [97]. Le premier modèle est une forme étendue du CTD en coude  $\beta$  de 650 Å et un *linker* de 250 Å [32]. Selon ce modèle, le CTD aurait la possibilité d'interagir avec toute la surface de l'ARNPII qui possède un diamètre de 150 Å. Le deuxième et le troisième modèles consistent en une structure en forme de spiral constitué d'un chevauchement de structure en coude  $\beta$ , soit deux par répétition (280 Å) ou une par répétition (100 Å), appelés respectivement spiral relâchée et compacte [93, 98]. Le dernier modèle est en forme de bobine avec une structure aléatoire [32]. Ce dernier a récemment obtenu des évidences biophysiques comme étant la structure du CTD chez *Drosophila melanogaster* avec un diamètre d'environ 100 Å. Puis, lorsqu'il y a phosphorylation du domaine, celui-ci augmente légèrement en diamètre (environ 120 Å) et en rigidité [68]. Ce relâchement permet une meilleure accessibilité pour des interactions avec d'autres protéines, comme le démontre la sensibilité aux protéases. De plus, cette nouvelle étude confirme des observations obtenues avec un peptide de deux répétitions où la Ser2 était phosphorylée et entraînait une étendue du peptide et une plus grande susceptibilité aux protéolyses [99, 100].

### 1.3.3 Caractéristique physico-chimique

Récemment, il a été démontré que le CTD peut participer à la formation et se localiser dans des gouttelettes issues d'une séparation de phase liquide-liquide [101-104]. Ces gouttelettes sont des compartiments sans membrane, les granules P et les corps Cajal en sont des exemples. Ce champ de recherche pour le CTD est relativement nouveau et ouvre de nouvelles possibilités dans la compréhension de la régulation transcriptionnelle. L'ARNPII est déjà connu pour former des agrégats dans des endroits précis du noyau durant la transcription [105, 106]. Ces gouttelettes sont formées par des interactions de protéines de faibles complexités avec des ARN qui impliquent une séparation de phase liquide-liquide [107]. Des études *in vitro* ont démontré que certains facteurs de transcription permettent le recrutement au promoteur de l'ARNPII via son CTD qui est un domaine à faible complexité [101-103]. Un CTD de cinq répétitions ne peut pas s'introduire dans les gouttelettes, mais si le nombre de répétitions est augmenté à 10 répétitions, il y a incorporation [101]. De plus, les séquences non consensus chez les mammifères sont mieux retenues dans ces gouttelettes que les séquences consensus. Il a été démontré récemment que la cycline T1 du facteur d'élongation *Positive transcription elongation factor B* (P-TEFb) contient un domaine riche en histidine et que celui-ci est nécessaire pour l'interaction entre la kinase de P-TEFb, CDK9, et le CTD afin de pouvoir le phosphoryler [65, 104]. De plus, ce domaine de la cycline peut former des gouttelettes et le CTD peut s'y loger. Cet effet est aussi amplifié par une pré-phosphorylation du CTD par TFIIF. Une autre étude *in vitro* démontre que le CTD à lui seul peut former des gouttelettes avec d'autres CTD, grâce à sa faible complexité protéique [103]. Puis, une fois phosphorylé, les différents CTD sont relâchés. Ces deux études proposent donc un rôle pour la formation de séparation de phase liquide-liquide dans la régulation de l'initiation transcriptionnelle.



### 1.3.4 Modifications post-traductionnelles

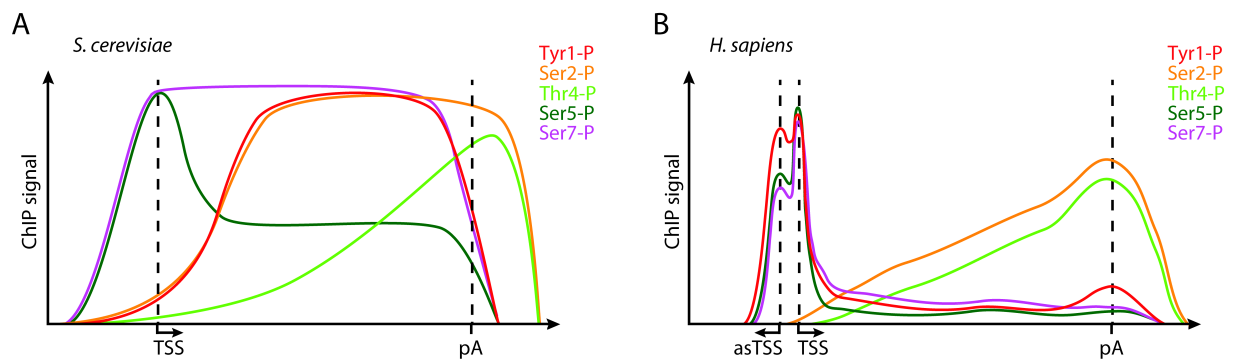
#### 1.3.4.1 Cycle de phosphorylation

Lors de la découverte du CTD en 1985, trois populations d'ARNPII, appelées Ilo, Ila et I Ib, avaient été découvertes en effectuant une séparation SDS-PAGE [60]. Ilo correspond à la forme hyperphosphorylée de l'ARNPII. Lorsque cette enzyme est traitée aux phosphatases, il y a déphosphorylation, et on observe la forme Ila [108]. Pour ce qui est de la dernière forme, elle correspond à une ARNPII sans CTD qui est très similaire au poids moléculaire de celle de l'ARNP d'*Escherichia coli*. Rapidement après cette découverte, il a été établi à l'aide d'expériences *in vitro* que l'ARNPII doit être non phosphorylé afin d'être capable d'initier la transcription suivie d'une phosphorylation lors de l'élongation [108-111]. D'ailleurs, il a été démontré que cette absence de phosphorylation lors de l'initiation est nécessaire à la formation du complexe de pré-initiation (PIC) afin que le CTD puisse être en interaction avec le complexe Médiateur [110-116]. Par la suite, à l'aide d'expériences génétiques et biochimiques, il a été démontré que le CTD peut être phosphorylé sur différents résidus et que ceux-ci impliquent des fonctions différentes [117, 118]. Un grand avancement est ensuite survenu avec l'élaboration d'anticorps reconnaissant les différentes marques de phosphorylation. Dans un premier temps, à l'aide de peptides du CTD et de la kinase CTDK-I, une première génération d'anticorps a été développée [119]. Les résultats obtenus ont permis de constater qu'il y a une phosphorylation du CTD durant l'élongation, contrairement à l'initiation [120, 121]. Par la suite, deux autres anticorps plus spécifiques ont été utilisés, H5 et H14, reconnaissant respectivement les marques de phosphorylation de la Ser2 et Ser5 [122-124]. Cela a permis de démontrer que le CTD subit des changements de phosphorylation dans ses sites durant des épisodes de stress cellulaire [122]. Puis, ces anticorps ont été utilisés avec la technique d'immunoprécipitation de la chromatine (ChIP) qui a permis d'obtenir les premiers résultats sur la localisation des marques de phosphorylation du CTD sur la chromatine chez *S. cerevisiae*. Ces résultats montrent que la Ser5 phosphorylée se situe principalement en 5' des gènes, dans la région du promoteur, tandis que la phosphorylation de la Ser2 se localise dans la région codante (ORF) correspondant à l'étape de l'élongation [125]. En se basant sur ces observations et sur le fait qu'il avait déjà été proposé à l'époque que le

CTD pouvait être en interaction avec le Médiateur, l'enzyme de coiffage (*capping enzyme*) et les facteurs de poly-adénylation, Stephen Buratowski proposa le code du CTD. Cette notion permit d'expliquer comment le CTD pouvait agir de plateforme, via une modulation de ses différentes phosphorylations, pour le recrutement spécifique de facteurs de transcription [126]. Enfin, avec l'élaboration d'autres anticorps spécifiques pour chacune des marques de phosphorylation et l'utilisation de la technique d'immunoprécipitation de la chromatine hybridée à des micropuces (ChIP-chip) ou couplée aux séquençage à haut-débit (ChIP-seq), le cycle de phosphorylation a été déterminé chez *S. cerevisiae*, *S. pombe* et l'humain [85, 87, 127-137].

Les études de localisation des marques de phosphorylation du CTD à l'échelle du génome chez la levure *S. cerevisiae* ont permis de démontrer qu'elles ont une distribution unique, très dynamique et indépendante de la longueur des gènes [129-131] (Figure 1.5A). La Ser5 est majoritairement phosphorylée dans la région promotrice puis elle subit une première déphosphorylation de 50 % environ 500 bp en aval du site d'initiation de la transcription (TSS) pour enfin être complètement déphosphorylée au site de poly-adénylation (poly(A)). La phosphorylation de la Ser7 est très similaire à celle de la Ser5, c'est-à-dire qu'elle débute et atteint un plateau au TSS. Contrairement à la Ser5, elle est stable durant la transcription et sa déphosphorylation n'est observée qu'à partir du site de poly(A). Il est toutefois important de noter que des études *in vitro* suggèrent que la Ser5-P est un prérequis pour la phosphorylation de la Ser7 [138]. Pour ce qui est de la Ser2, comme entendu, sa phosphorylation est progressive et atteint un plateau à environ 1 kb du TSS. Elle est ensuite déphosphorylée après le site de poly(A). Quant au patron de phosphorylation de la Tyr1, elle est identique à celle de la Ser2 à une différence près [128]. Le résidu est déphosphorylé en amont du site de poly(A), contrairement à celle de la Ser2 qui se situe en aval. Enfin, la phosphorylation de la Thr4 est plutôt controversée. Notre laboratoire n'a pas été en mesure de détecter la phosphorylation de ce résidu. Toutefois, le groupe de Patrick Cramer observe une distribution faible dans l'ORF, soit entre le TSS et le site de poly(A) [128]. De plus, l'anticorps reconnaissant la marque Thr4-P avait été utilisé en immunobuvardage de type Western sur des extraits cellulaires de *S. cerevisiae* et un signal a été détecté [85]. Cependant, ce signal est très faible par rapport aux autres marques et à ce qu'il est détecté de la Thr4-P chez *S. pombe* et les mammifères. Ainsi, il

a été suggéré que cette phosphorylation est absente chez *S. cerevisiae*. Toutefois, au courant des dernières années, deux autres laboratoires ont utilisé deux techniques différentes utilisant le même anticorps pour détecter la Thr4-P à l'échelle du génome chez *S. cerevisiae*, soit le mCRAC et le ChIP-nexus [133, 134]. Dans un premier temps, les résultats obtenus pour le mCRAC montrent une distribution similaire à celle obtenue par Mayer *et al.* Pour ce qui est de la deuxième étude utilisant la technique ChIP-nexus, les auteurs détectent une faible quantité de Thr4-P durant l'élongation suivie d'une augmentation qui atteint son maximum en aval du site poly(A).



**Figure 1.5 Cycle de la phosphorylation du CTD de l'ARNPII.**

A) Patron de phosphorylation du CTD déterminé en immunoprécipitation de la chromatine (ChIP) chez *S. cerevisiae*. La phosphorylation de la Ser5 et de la Ser7 est détectée au site d'initiation de la transcription (TSS). Puis à 500 bp, 50 % de la Ser5-P est déphosphorylée et une augmentation de la phosphorylation de la Tyr1 et de la Ser2 est détectée pour atteindre un plateau maximum à 1000 bp. La phosphorylation de la Thr4 est, quant à elle, plus tardive et atteint son maximum près du site de poly-adélation (pA). Au même site, une déphosphorylation de la Tyr1, de la Ser5 et de la Ser7 a lieu. Enfin, la déphosphorylation de la Ser2 et Thr4 est effectuée en aval du site pA. B) Patron de phosphorylation du CTD déterminé en ChIP chez l'humain. Le cycle de phosphorylation est similaire à celui chez la levure. Toutefois, la Tyr1-P se localise principalement au TSS. La Tyr1-P, la Ser5-P et la Ser7-P sont aussi détectées au TSS anti-sense (asTSS). Figure adaptée de Jeronimo *et al.* 2016 [139].

Les résultats obtenus chez *S. pombe* sont pratiquement identiques à ceux de *S. cerevisiae* [137]. En effet, les patrons de phosphorylation de la Ser2, la Ser5 et la Ser7 sont similaires entre les deux levures. La seule différence est l'endroit où la Tyr1-P est déphosphorylée. Chez *S. pombe*, on observe une diminution de la phosphorylation en aval du site poly(A), contrairement à *S. cerevisiae* où c'est en amont.

Ces résultats pour le cycle de phosphorylation chez *S. cerevisiae* sont aussi très similaires à ceux observés chez les mammifères [85, 87, 127, 135, 136]. La différence majeure est toutefois pour la Tyr1. En effet, contrairement à la levure, la distribution de la Tyr1-P se situe en 5' de façon similaire à la Ser5-P et à la Ser7-P. Il est aussi possible d'observer une phosphorylation de ces trois résidus aux promoteurs divergents (Figure 1.5B) et aux *enhancers*, endroit où la Tyr1-P est prédominante. Il a été suggéré par les auteurs que cette phosphorylation est une marque de cette région régulatrice. Ils ont également détecté une forte corrélation entre la Tyr1-P et les sites de pause de l'ARNPII, ce qui a aussi été validé avec un mutant de la Tyr1 [87, 88]. De plus, sa phosphorylation a été proposée pour être impliquée dans la stabilité du domaine, sans quoi il y a dégradation du CTD par le protéasome [136].

Donc, tous ces résultats ont été possibles grâce aux développements d'anticorps spécifiques pour chacune des marques de phosphorylation du CTD [123, 124, 127, 128]. Cependant, ceux-ci ont plusieurs limitations. Premièrement, la spécificité d'un anticorps peut être affectée par la phosphorylation d'un résidu adjacent [85, 128]. Dans ce cas, une perte de phosphorylation n'est pas obligatoirement réelle, de même qu'une augmentation de la phosphorylation. Ils ne reconnaissent que les séquences consensus, ce qui représente une limite importante pour l'étude du CTD de la drosophile et de la séquence non-consensus du CTD de l'humain. Une réactivité croisée des anticorps est aussi possible. Par exemple, l'efficacité de détection de la Ser2-P à l'aide de l'anticorps H5 est augmentée lorsqu'il y a phosphorylation de la Ser5 sur le même peptide [127]. De plus, pour une même marque, comme celle de la Ser2, deux anticorps différents peuvent exister et donner des résultats différents. Ceci démontre la limite de ces outils à détecter certaines variantes de phosphorylations spécifiques aux détriments d'autres [130]. Dans le même ordre idée, l'utilisation d'anticorps ne peut distinguer à quel endroit la phosphorylation d'un résidu à lieu

sur les différentes sections proximales ou distales du CTD, car on sait que la fonction de ces deux parties est différente [72].

Afin de répondre à cette dernière limite, deux laboratoires ont employé un système *in vivo* utilisant différents mutants du CTD, à la fois chez *S. cerevisiae* et l'humain, couplé à la spectrométrie de masse [140, 141]. La stratégie consiste à remplacer certains résidus de la Ser7 par des Arg ou des Lys afin de rendre le CTD sensible à la digestion à la trypsine. À l'aide des différents peptides obtenus, il a donc été possible de détecter les marques sur chacune des 26 ou 52 répétitions. Les résultats obtenus donnent lieu à des observations étonnantes. En effet, ils ont constaté que la phosphorylation avait lieu sur tout le CTD et non seulement dans la région proximale ou distale. Ils ont aussi remarqué que la plupart des heptapeptides ne sont pas phosphorylés et que ceux qui le sont ne comportent qu'une seule modification. Les marques Ser2-P et Ser5-P sont les plus abondantes, contrairement à la Tyr1-P et à la Thr4-P qui sont très faibles, voire pratiquement absentes. Ce qui est étonnant est la faible proportion de Tyr1 phosphorylée, car celle-ci a été proposée pour être requise afin de recruter des facteurs d'élongation et de bloquer le recrutement précoce de facteurs de terminaison [128, 142]. Pour ce qui est de la Ser7, son niveau de phosphorylation est faible, mais peut être expliqué par le fait que ce résidu a été muté dans la moitié des heptapeptides. Ces études ont toutefois certaines limitations. Les différentes mutations des Ser7 en Arg et en Lys sont de loin similaires au CTD sauvage. Elles peuvent aussi avoir des impacts sur la structure du domaine et, par le fait même, affecter sa reconnaissance et l'efficacité des différentes kinases et phosphatases. Enfin, les marques de phosphorylation correspondent à l'ARNPII totale, ce qui peut avoir eu comme conséquence de sous-estimer des populations avec des patrons différents.

#### **1.3.4.1.1 Kinases et phosphates**

Ces différents patrons de phosphorylation du CTD sont rendus possibles grâce à différentes kinases et phosphatases (Table 1.1). Pour commencer, lors de la formation du PIC, l'ARNPII est recrutée sous sa forme non-phosphorylée [108-111]. Par la suite, la kinase du

complexe TFIIH, soit Kin28 chez *S. cerevisiae* ou Cdk7 chez l'humain, phosphoryle spécifiquement le résidu de la Ser5 du CTD avant le début de l'initiation [125, 130, 138, 143, 144]. Cette première modification permet la phosphorylation de la Ser7 et le relâchement de l'ARNPII au promoteur et ainsi le début de l'initiation [138, 145-151]. Chez les métazoaires, un évènement généralisé de pause majeur a lieu dans la région proximale du promoteur qui correspond à environ 20-100 bp après l'initiation [152-156]. Cet élément de régulation est causé par deux facteurs négatifs de l'élongation, c'est-à-dire le *negative elongation factor* (NELF) et le *DRB-sensitivity-inducing factor* (DSIF) [157-159]. Il est important de noter que le facteur NELF n'est pas présent chez la levure, ce qui explique l'absence de pause en 5' [160]. Donc afin de sortir de cette impasse, le facteur d'élongation P-TEFb, qui est constitué chez l'humain d'une cycline T et d'une kinase Cdk9, phosphoryle les facteurs NELF et DSIF [161-164]. Cela a pour effet de permettre le relâchement de NELF et de rendre DSIF positif à l'élongation [165, 166]. De plus, P-TEFb effectue la phosphorylation de la Ser2 du CTD qui est stimulée lorsque la Ser7 est préalablement phosphorylée [65, 167, 168]. Il a aussi été démontré que cette kinase peut avoir la Ser5 et la Ser7 comme substrat lors d'expérience *in vitro* [167]. Pour ce qui est de la phosphorylation de la Ser2 chez *S. cerevisiae*, deux kinases permettent cette modification, soit Ctk1 (complexe CTDK-I) et Bur1 (homologue de P-TEFb) [119, 130, 164, 169-173]. Des évidences démontrent que Ctk1 est la kinase principale de la Ser2 et que celle-ci requiert préalablement une phosphorylation de la Ser2 par Bur1 pour être stimulée [130, 172]. Contrairement à TFIIH qui se localise seulement au promoteur, les kinases de la Ser2 se localisent dans l'ORF [80, 173, 174]. Il a aussi été démontré que ces différentes enzymes sont stimulées par l'activité de phosphorylation de TFIIH. En effet, la Ser5-P stimule la phosphorylation de la Ser2 via Ctk1 chez la levure et TFIIH permet le recrutement de P-TEFb [172, 175-177]. L'homologue de Ctk1 chez l'humain est Cdk12 et elle permet la phosphorylation *in vitro* des trois sérines du CTD [169, 178]. Toutefois, son effet *in vivo* est plutôt modeste pour la Ser2 [169, 179-181]. Une observation intéressante avec ces différentes kinases de la Ser2 a été effectuée chez la drosophile. En effet, contrairement à l'humain qui a comme kinase principale P-TEFb, celle de la mouche est Cdk12 [169]. Plus précisément, P-TEFb est requis pour une phosphorylation du résidu en 5' et Cdk12 lors de l'élongation [169]. Une autre kinase a aussi été caractérisée pour la Ser2 chez l'humain [182]. Appelée Brd4, elle permet l'obtention d'un variant de la Ser2-P différent de celui obtenu par

P-TEFb [182]. En effet, la phosphorylation obtenue par Brd4 est reconnue par l'anticorps 3E10, tandis que celle de P-TEFb est reconnue par H5 [182, 183]. Étonnamment, il a été démontré que Brd4 est nécessaire pour le recrutement de P-TEFb [184, 185]. Comme TFIIH est aussi nécessaire au recrutement de P-TEFb, il serait intéressant de savoir si cela est rendu possible par le recrutement de SAGA par la Ser5-P du CTD qui permettrait par la suite un recrutement de Brd4, via son domaine bromo, et ainsi ultimement P-TEFb. Ces différentes kinases de la Ser2 se localisent toutes dans la région codante des gènes, sauf Brd4 qui se localise aux promoteurs ou aux *enhancers* [80, 173, 174, 186]. Pour ce qui est des résidus Tyr1 et Thr4, des kinases ont été identifiées chez l'humain. La kinase de la Tyr1, c-Abl, qui ne possède aucun homologue chez la levure, semble être plutôt spécifique à certains gènes et requiert des conditions précises [187-191]. Des expériences ont tenté de savoir si cette kinase pouvait phosphoryler la Tyr1 du CTD de la levure, mais comme la kinase semble nécessiter la dernière répétition du domaine des mammifères, aucune modification n'a été détectée [190]. Récemment, il a été démontré que, lors de situations de stress spécifiques, la MAP kinase Slk2 pouvait effectuer une phosphorylation de la Tyr1 [192]. Enfin, *Polo-like kinase 3* (Plk3) a été proposée pour être la kinase de la Thr4 et aucune kinase homologue n'a été découverte jusqu'à maintenant chez la levure [85]. Plusieurs autres kinases ont aussi été découvertes pour le CTD, comme ERK1/MAPK1, DYRK1A et CDK13, mais leurs rôles sont beaucoup moins bien compris.

Ces différentes marques doivent ensuite être supprimées afin de compléter le cycle et d'obtenir une déphosphorylation totale de l'ARNPII lors de la terminaison pour permettre un recyclage de l'enzyme. Pour y arriver, ceci nécessite différentes phosphatases. Quatre enzymes majeures ont été découvertes jusqu'à maintenant : Rtr1, Fcp1, Ssu72 et Glc7. Rtr1 est la première à entrer en jeu lors de la transcription. Elle permet la déphosphorylation de la Ser5 durant les premières 500 bp de la transcription [193, 194]. Toutefois, aucun mécanisme n'a été découvert pour expliquer comment la phosphatase déphosphoryle seulement 50 % des marques. Chez l'humain, l'homologue de Rtr1, appelé RPAP2, est aussi connu pour être spécifique à la Ser5 [194, 195]. L'enzyme est recrutée sur le CTD à l'aide de deux partenaires d'interaction (RPRD1A et RPRD1B) qui ont comme substrats la Ser2-P et/ou la Ser7-P [195]. Cela permet ensuite la déphosphorylation de la Ser5 se situant entre les résidus Ser2-P et/ou la

Ser7-P. Toutefois, les deux partenaires de RPAP2 n'existent pas chez la levure. Il est intéressant de noter que Rtr1 est dépendant de Ctk1 pour son interaction avec le CTD, ce qui permet d'imaginer un modèle où la Ser2-P est nécessaire à son recrutement pour ensuite occasionner la déphosphorylation de la Ser5 adjacente [196]. Pour ce qui est de Fcp1, elle est la première phosphatase découverte du CTD [197, 198]. Elle a pour substrat la Ser2 *in vivo*, et ce, même si elle peut avoir les trois sérines du CTD comme substrats *in vitro* [173, 199-201]. À elle seule, elle n'est pas capable d'effectuer une déphosphorylation totale de la Ser2 après la terminaison comme le démontre des expériences *in vitro*. Pour que la phosphatase puisse être capable d'effectuer une déphosphorylation totale, Fcp1 requiert l'aide de la phosphatase Ssu72 [80, 130]. Cette dernière fait partie du module phosphatase du complexe de terminaison *cleavage and polyadenylation factors* (CPF) [202, 203]. Elle permet la déphosphorylation de la Ser5 et de la Ser7 au site poly(A) [80, 130, 204, 205]. Comment la phosphatase fait-elle la différence entre ces deux résidus? La réponse est dans le type d'isomérisation de la proline 6 du CTD. En effet, lorsque le CTD est en conformation *trans*, Ssu72 est spécifique à la Ser7 et lorsque le domaine est en *cis*, elle est spécifique à la Ser5 [206-208]. Ceci est rendu possible par l'isomérase Ess1 qui a été démontré pour stimuler l'activité de Ssu72 pour ces deux résidus du CTD de façon *in vitro* [130, 208, 209]. Finalement, Glc7, une enzyme du module phosphatase du complexe CPF, permet la déphosphorylation de la Tyr1 au site poly(A) chez *S. cerevisiae*. Cette déphosphorylation en amont de celle de la Ser2 du CTD a été proposée pour permettre le recrutement du complexe CPF, impliqué dans la terminaison. Récemment, il a été démontré que l'équivalent chez *S. pombe*, Dis2, permet plutôt la déphosphorylation de la Thr4 du CTD [210]. Toutefois, aucune évidence ne démontre que l'homologue chez l'humain, PP1, est responsable d'une de ces déphosphorylations.



**Table 1.1 Kinases et Phosphatases du CTD ayant des évidences *in vivo*.**

	<i>S. cerevisiae</i>	<i>Homo sapiens</i>	<i>S. pombe</i>	<i>D. melanogaster</i>	Complexe protéique	Résidu spécifique
Kinases	Kin28*	Cdk7*	Mcs6	Cdk7	TFIIH	Ser5 et Ser7
	Bur1*	Cdk9*	Cdk9	Cdk9	P-TEFb	Ser2
	Ctk1*	Cdk12*	Lsk1	Cdk12*	CTDK-I	Ser2
	-	c-Abl*	-	-	-	Tyr1
	Slr2*	Erk1/2	Pmk1	Erk	-	Tyr1
	-	Plk3*	-	-	-	Thr4
Phosphatases	Rtr1*	RPAP2*	Rtr1	CG34183	-	Ser5
	Fcp1*	Fcp1	Fcp1	Fcp1	-	Ser2
	Ssu72*	Ssu72	Ssu72	Ssu72	CPF	Ser5 et Ser7
	Glc7*	PP1	Dis2	PP1	CPF	Tyr1
	Glc7	PP1	Dis2*	PP1	CPF	Thr4

L'astérisque indique dans quel organisme l'enzyme a été caractérisée pour avoir le résidu du CTD comme substrat.

### 1.3.4.2 Autres

Même si la phosphorylation est la modification post-traductionnelle principale du CTD, d'autres sont aussi possibles, majoritairement chez les eucaryotes supérieurs. L'une d'elles est la glycosylation des trois sérines et de la thréonine. Connue depuis les années 1980, cette modification n'est pas présente chez la levure. Les deux enzymes permettant d'ajouter et d'enlever ce monosaccharide ont été identifiées chez l'humain : O-GlcNAc transférase et N-acétylglucosamidase, respectivement. Comme la glycosylation a lieu durant la formation du PIC, il a été proposé qu'elle est impliquée dans la prévention de la phosphorylation du CTD [211-213]. Il a aussi été démontré que d'autres modifications de résidus spécifiques se localisent sur les séquences non-consensus. En effet, la méthylation de l'Arg1810 par CARM1 est impliquée dans la régulation transcriptionnelle des petits ARN nucléaires (snRNA) et petits ARN nucléolaires (snoRNA) [214]. Il a aussi été démontré que cette modification, cette fois-

ci effectué par PRMT5, est nécessaire pour la résolution d'hybride ADN-ARN, appelé *R-loop*, et que cela est requise pour permettre la terminaison transcriptionnelle des gènes codants chez les mammifères [215]. Les Lys7 du CTD chez l'humain et la souris peut aussi être méthylée ou acétylée. La méthylation est proposée pour être impliquée dans la régulation négative de l'expression des gènes [216, 217]. Quant à l'acétylation, elle est impliquée dans l'élongation et dans l'induction de certains gènes de croissance [216-218]. Enfin, une dernière modification connue du CTD est l'ubiquitylation de résidus Lys chez la souris qui est impliquée dans la dégradation de Rpb1 [219].

### 1.3.5 Protéines liant les différentes formes phosphorylées

Au fil des ans, plusieurs études ont démontré que les différentes phosphorylations du CTD servaient de plateforme pour le recrutement de facteurs impliqués dans la maturation des pré-ARNm, permettant ainsi un couplage avec la transcription [220]. De plus, de nombreuses protéines et enzymes impliquées dans les étapes de la transcription ont aussi été identifiées [65]. Pour certaines d'entre elles, les résidus phosphorylés du CTD nécessaire à leurs interactions ont été identifiés. Toutefois, seulement un nombre restreint de ces partenaires a vu ses domaines d'interaction caractérisés.

Pour commencer, lors de la formation du PIC, il a été démontré que le CTD est en interaction directe avec deux modules du Médiateur, soit le *head* et le *middle* [114, 115, 221]. L'interaction implique quatre heptapeptides du CTD d'une longueur de 73 Å [115]. Le domaine de l'ARNPII doit être non-phosphorylé et posséder une conformation étendue avec une structure en coude  $\beta$  à son centre [100, 115, 221, 222]. Cette interaction avec le Médiateur est nécessaire afin de positionner le CTD pour que la kinase de TFIIH puisse effectuer une phosphorylation de la Ser5 [114, 223]. Pour que cela soit possible, il a été suggéré que le Médiateur subit un changement de conformation pour rendre le CTD accessible à la kinase [223]. Une fois la Ser5 phosphorylée, le Médiateur relâche le CTD et il y a début de la transcription.

Dès que les ARN naissants sortent du canal de sortie, le complexe de fixation de la coiffe permet l'ajout d'un m<sup>7</sup>GpppN afin de les protéger de la dégradation par les exonucléases [224, 225]. Les premières évidences démontraient que le recrutement du complexe nécessitait un CTD phosphorylé pour la Ser5 [226-230]. Par la suite, des expériences de cristallographie de la guanylyltransférase Cgt1 (*C. albicans*) et un peptide contenant quatre heptapeptides du CTD, où les Ser5 étaient phosphorylées, ont été effectuées [231]. Les résultats démontrent que le CTD est en conformation étendue *β-like* et que l'enzyme possède deux régions d'interaction qui impliquent deux Ser5-P, soit un sur le premier heptapeptide et un sur le troisième. De plus, la Tyr1 est aussi importante pour effectuer un pont hydrogène et une interaction hydrophobique avec l'enzyme. Étonnamment, des résultats différents ont été obtenus avec la levure *S. pombe*. Avec cet organisme, l'homologue de Cgt1, Pce1, ne possède qu'un site d'interaction avec le CTD nécessitant huit acides aminés ayant deux Ser5-P aux extrémités qui permettent l'interaction [232]. Quant à la Tyr1, elle est encore une fois impliquée dans des interactions hydrophobiques et dans la formation de ponts hydrogènes. Enfin, une structure a aussi été déterminée pour l'homologue chez la souris, Mce1 [233]. La conformation du CTD est la même que celle observée chez les deux levures et l'interaction requiert aussi la Tyr1, mais elle nécessite cette fois-ci seulement une répétition de l'heptapeptide où la Ser2 et la Ser5 sont phosphorylées. Même s'il y a présence de la Ser2-P, aucune interaction n'est observée avec ce résidu modifié.

Le domaine d'interaction au CTD le plus étudié est nommé CID pour *C-terminal interaction domain*. Ce dernier se retrouve fréquemment dans les facteurs de terminaison chez *S. cerevisiae* : Pcf11, Rtt103 et Nrd1 [94, 98, 234-241]. Il est aussi présent dans d'autres protéines chez les mammifères : SCAF4, SCAF8, RPRD1A et RPRD1B [242-244]. La structure du CID est composée de 140 acides aminés en N-terminal et elle est généralement formée de huit hélices possédant un centre hydrophobique [98]. L'interaction implique toutefois seulement trois hélices et 8 à 11 résidus du CTD. Lorsque ce dernier est en interaction avec le CID, il y a un changement de conformation pour une structure en coude  $\beta$ , où l'isomérisation des prolines est en *trans*, sauf pour l'interaction avec Nrd1 qui est en *cis* [245]. Le CID de Pcf11 se caractérise par le fait qu'il ne forme pas d'interaction directe avec

la Ser2-P du CTD, contrairement à Rtt103 [98, 235]. Une perte de phosphorylation de la Ser2 par l'entremise d'un mutant *ctk1*Δ entraîne une perte de recrutement du facteur Rtt103, ce qui n'est pas le cas avec Pcf11, où on observe un effet modeste [234, 240]. Dans ce dernier cas, la Ser2-P du CTD semble plutôt être impliquée dans la stabilisation de la conformation en coude β du CTD, ce qui explique l'implication partielle du résidu dans le recrutement de Pcf11 [94, 98]. Récemment, il a été démontré que le CID de Rtt103 est en interaction directe avec le résidu de la Thr4 du CTD [246, 247]. En fait, deux possibilités d'interaction ont été observées. Dans un premier temps, lorsque le CTD est phosphorylé à sa Ser2, la Thr4 est impliquée dans un pont hydrogène avec la Ser2-P pour stabiliser la structure en coude β. Puis, lorsque la Ser2 n'est pas phosphorylée, la Thr4-P est nécessaire pour une interaction via des ponts hydrogènes avec le CID. Des interactions importantes sont aussi visibles pour la Tyr1 et la Pro3 du CTD. Puis, lorsque la Tyr1 est phosphorylée, il y a un encombrement stérique qui empêche l'interaction. De plus, lorsque l'arginine du CID identifiée pour être en interaction directe avec la Tyr1 du CTD est mutée, un défaut de croissance est observé [248]. Ainsi, selon ces observations et d'autres évidences qui seront soulevées dans la dernière section de l'introduction, il a été proposé que la phosphorylation de la Tyr1 joue un rôle clé dans la terminaison. Celui-ci serait de bloquer un recrutement précoce des facteurs de terminaison lors de l'élongation [128, 142]. Pour ce qui est de la protéine SCAF8, elle est en interaction avec le CTD via la Ser2-P [242, 243]. Finalement, les protéines RPRD1A et RPRD1B, impliquées dans le recrutement de la phosphatase RPAP2, contiennent aussi des domaines CID [195, 244]. L'interaction nécessite la Ser2-P et/ou la Ser7-P de chaque côté de la Ser5 du CTD à déphosphoryler [195].

Plusieurs autres domaines d'interaction au CTD ont aussi été identifiés : SH2, SR1, WW, FF, RRM et Tudor. Le motif *Src homology 2* (SH2), d'une longueur d'environ 100 acides aminés, est connu pour se lier à une tyrosine phosphorylée [249]. La protéine de remodelage des nucléosomes Spt6 possède deux de ces domaines [250]. Utilisant différents peptides phosphorylés, des études ont démontré que le domaine SH2 de Spt6 peut se lier à la Ser2-P et à la Tyr1-P, mais l'interaction avec cette dernière possède une plus grande affinité [251, 252]. Récemment, il a toutefois été démontré que Spt6 effectue principalement des

interactions avec le *linker* du CTD [253, 254]. Le domaine SR1 que l'on retrouve dans la méthyltransférase Set2 permet une interaction avec deux heptapeptides, où les deux Ser2 et Ser5 du CTD sont phosphorylées [255, 256]. Il est constitué de 100 acides aminés formant trois hélices [256]. Le domaine WW est aussi connu pour permettre une interaction avec le CTD [97]. Comportant 40 acides aminés et deux tryptophanes séparés par environ 20-22 résidus, ce domaine possède une conformation de trois feuillets  $\beta$  anti-parallèles [257, 258]. Il est connu pour permettre une interaction avec des sérines ou des thréonines phosphorylées et des peptides riches en proline [259, 260]. L'isomérase Ess1 (Pin1 chez l'humain), qui est en interaction avec la Ser5-P du CTD, contient un de ces domaines en N-terminal [261-263]. Une structure pour le domaine WW de Pin1 a été déterminée en cristallographie avec ou sans un peptide du CTD phosphorylé pour la Ser2 et la Ser5 [264, 265]. C'est toutefois seulement avec la Ser5-P et les résidus prolines en conformation *trans* du CTD qu'il y a une interaction directe avec le domaine. Quant au domaine FF, il contient 60 acides aminés, deux phénylalanines et est souvent accompagné du domaine WW. Il est connu pour être en interaction avec des protéines phosphorylées [266]. En conformation de bouquet de trois hélices, on le retrouve chez différents facteurs d'épissage qui sont en interaction avec la Ser2-P et Ser5-P du CTD, dont Prp40 [267]. Plusieurs de ces domaines sur une même protéine peuvent être impliqués dans une interaction avec le CTD [267, 268]. Le domaine *RNA recognition motif* (RRM) est un motif qui permet l'interaction entre une protéine et un ARN. Celui de la protéine Yra1, impliqué dans l'exportation des ARN hors du noyau, peut être à la fois en interaction avec l'ARN et le CTD [269]. En effet, cette protéine est en interaction directe avec la Ser2-P et la Ser5-P du CTD, tandis que le contact avec l'ARN est beaucoup plus faible. Enfin, le domaine Tudor de Tdrd3 chez l'humain est en interaction avec les répétitions 31 et 32 du CTD lorsque l'Arg1810 est diméthylée de façon asymétrique [270]. Une interaction de SMN par son domaine Tudor avec l'Arg1810me2 du CTD a aussi été démontrée [215]. Aucune interaction avec l'arginine n'est toutefois observée, les contacts sont plutôt avec les prolines. Fait intéressant, la Tyr1 n'est pas impliquée dans aucune interaction avec ce motif, contrairement aux autres domaines détaillés précédemment, où le résidu est toujours impliqué dans des ponts hydrogènes ou des interactions hydrophobes.

Les protéines de faible complexité sont aussi connues pour former des interactions avec le CTD. Comme mentionné dans la section 1.3.3., les protéines P-TEFb (cycline T1), FUS et TAF15 possèdent ces caractéristiques qui permettent de former avec le CTD des gouttelettes qui se résument en une séparation de phase liquide-liquide [101, 102, 104]. Celles-ci sont proposées pour jouer un rôle dans la régulation de l'initiation. Jusqu'à maintenant, on en connaît peu sur ces types d'interaction impliquant le CTD et les facteurs de transcription, mais il ne fait aucun doute que de prochaines études permettront une meilleure compréhension de ce phénomène.

## 1.4 Transcription : Initiation et élongation

### 1.4.1 Initiation

La transcription est régulée par deux régions sur l'ADN : l'activateur (*enhancer*) et le promoteur. Chez la levure, l'activer est nommé *upstream activator site* (UAS). Il se localise en amont du promoteur dans une région dépourvue de nucléosome, appelée *nucleosome free region* (NFR), tandis que chez les mammifères, l'activateur peut se situer en amont ou en aval du gène [271-273]. Cette région correspond à l'endroit où les facteurs de transcription se lient à l'ADN pour permettre la formation du PIC au promoteur. Pour ce qui est de ce dernier, il contient différents motifs pour le recrutement de facteurs spécifiques chez les métazoaires : la boîte TATA, l'initiateur, le motif à dix éléments (MTE), l'élément du promoteur en aval (DPE) et l'élément central en aval (DCE) [274, 275]. Toutefois, seulement la boîte TATA est présente chez la levure, et ce, dans 20-30 % des promoteurs [276, 277]. Ce motif se situe à environ 30 bp en amont du TSS chez les métazoaires, tandis que chez la levure, il est à environ 50-120 bp [278, 279].

Chez la levure, les gènes sont divisés en deux classes, c'est-à-dire ceux de croissance et ceux répondant à des situations de stress [280]. La première catégorie, aussi appelée les gènes *housekeeping*, est exprimée de façon constante dans la cellule, mais diminue lors d'un épisode de stress au profit d'une augmentation de l'expression de la deuxième catégorie de gènes. Les deux classes sont contrôlées par différents facteurs de transcription. En effet, les

gènes de croissance dépendent du facteur général de la transcription (GTF) TFIID, tandis que les gènes de stress sont régulés par le complexe SAGA (Spt-Ada-Gcn5 acétyltransférase) [281, 282]. De plus, les gènes de croissance possèdent une région NRF bien définie et ne contiennent pas de boîte TATA dans leurs promoteurs. Pour ce qui est des gènes de stress, ils possèdent une boîte TATA à leurs promoteurs et leurs régions NRF sont moins bien définies [280]. Il est bien établi dans la communauté scientifique que les gènes contrôlés par SAGA et par TFIID sont de 10 % et de 90 %, respectivement [281, 282]. Toutefois, cette vision binaire du contrôle de la transcription par ces deux facteurs a été révisée tout récemment [283, 284]. En effet, le laboratoire de Steven Hahn a démontré que tous les gènes avec ou sans boîte TATA dans leurs promoteurs étaient dépendants de TFIID. Puis, le laboratoire de Laszlo Tora en est venu à la même conclusion avec SAGA. Ces études démontrent aussi que SAGA se situe préférentiellement aux UAS et TFIID aux TSS. La dépendance des gènes pour un facteur ou l'autre varie selon les conditions de croissance et de stress. En d'autres mots, les deux articles démontrent que tous les promoteurs dépendent à la fois des deux facteurs et qu'ils effectuent une fonction complémentaire pour permettre l'initiation. Mais une question demeure : pourquoi les premières études ont démontré que 10 % des gènes étaient régulés par SAGA et 90 % par TFIID? Les différents types d'expériences effectués expliquent cette différence. En effet, initialement, pour identifier les gènes dépendant d'un ou l'autre de ces facteurs, les groupes de recherche utilisaient le CHIP pour déterminer la distribution des deux facteurs. Toutefois, cette technique utilise le formaldéhyde, un agent de fixation (*crosslinking*), qui peut amener des biais. Pour les expériences phénotypiques, ils utilisaient des protéines mutées pour les rendre sensible à certaines températures afin de les inactiver avec un choc thermique. Ce traitement peut toutefois avoir aussi des effets sur la transcription. Finalement, dans les études initiales, l'ARN total de la cellule était analysé après l'inactivation des protéines d'intérêt. Cependant, une diminution de TFIID permet de stabiliser l'ARN total et cache l'effet sur les ARN nouvellement synthétisés qui reflètent la réalité, par exemple. Pour éviter ces limitations, les auteurs de ces deux études utilisent une technique indépendante du formaldéhyde (ChEC), des techniques ciblant seulement l'ARN naissant et la technique *anchor-away* ou un système de dégradation inductible à l'auxine pour inactiver les protéines.

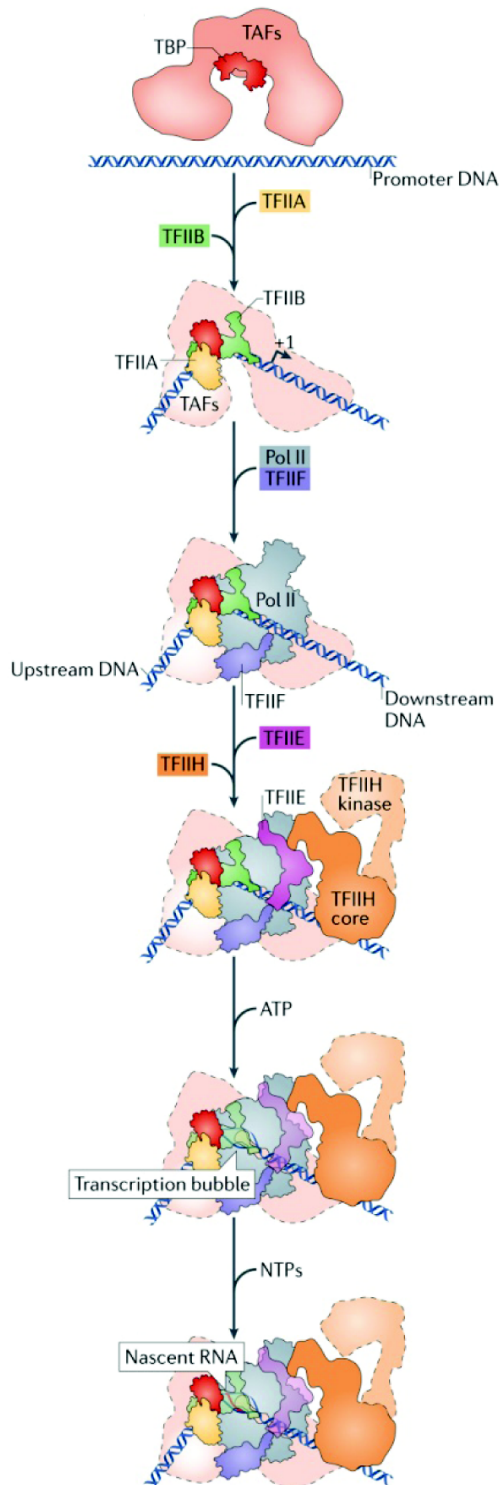
SAGA a été initialement identifié comme étant une acétyltransférase d'histone (HAT) à l'aide de sa sous-unité Gcn5 [285]. Il est important de noter que même s'il a été démontré que Gcn5 permet l'activation de la transcription, il est aussi impliqué dans la répression de certains gènes chez la levure [286, 287]. Ce complexe est très bien conservé dans l'évolution [288]. Chez les mammifères, il joue aussi un rôle après la formation du PIC [289]. SAGA est constitué de 21 sous-unités et forment cinq modules chez la levure : le corps du complexe, le HAT, la déubiquitineuse, l'activateur de liaison et les facteurs associés au TBP (TAF) [290]. Le corps a comme rôle de maintenir l'intégrité du complexe [285, 291-293]. Le HAT catalyse l'acétylation de l'histone H3, ce qui provoque une répulsion des charges des nucléosomes et permet l'ouverture de la chromatine [291, 294-296]. Le module déubiquitineuse contient une enzyme Ubp8 impliquée dans la déubiquitination de l'histone H2B123ub1 [297-301]. Cette réaction permet, entre autres, le recrutement de la kinase Ctk1 du CTD durant l'élongation [302]. L'activateur de liaison contient la protéine Tra1 nécessaire à l'interaction et au recrutement du complexe SAGA par un activateur de la transcription [303, 304]. Enfin, le dernier module contient Spt3 et Spt8 nécessaires à l'interaction entre SAGA et TBP [291, 305, 306].

Pour ce qui est de TFIID, il est constitué de quatorze TAF et de TBP chez la levure [307]. Ce GTF est aussi conservé chez les métazoaires où l'on retrouve TBP et treize des TAF de la levure. De plus, chez les métazoaires, plusieurs variantes de TFIID existent et elles sont impliquées dans l'expression de gènes spécifiques et dans des types cellulaires différents [308, 309]. Son rôle est de reconnaître, à l'aide de ses différents TAF et de TBP, le promoteur via les différents motifs [307, 310, 311]. Les séquences A/T riche de la boîte TATA permettent l'obtention d'une surface hydrophobe ayant une affinité avec le TBP [312]. Cette interaction implique très peu de contacts et elle permet de plier l'ADN dans un angle de 90 degrés qui sert ensuite de plateforme pour la formation du PIC [312, 313]. De plus, TFIID contient différents TAF impliqués dans la régulation de l'interaction TBP-ADN. En effet, le domaine TAND1 de Taf1 bloque la liaison du TBP avec la boîte TATA et le domaine TAND2 est un régulateur qui est en compétition avec TFIIA pour un contact avec le TBP [314-316]. En plus de Taf1, Taf11 et Taf13 compétitionnent avec la boîte TATA pour une liaison avec le TBP [317]. Ces différentes interactions permettent ainsi de prévenir une liaison non spécifique entre le TBP et



l'ADN. Une fois le TBP associé au promoteur, les différents GTF et le Médiateur sont recrutés pour former le PIC qui recouvre environ 60 bp de l'ADN [318] (Figure 1.6). Ces différents facteurs ont pour but de recruter l'ARNPII, d'ouvrir l'ADN pour permettre la transcription et de permettre la dissociation du promoteur.

À la suite du recrutement de TBP, les GTF TFIIA et TFIIB sont recrutés de part et d'autre de celui-ci [320]. TFIIA est composé de deux sous-unités chez la levure et de trois chez l'humain [321, 322]. Ce facteur permet de stabiliser l'interaction entre le TBP et l'ADN [320]. Il se lie au complexe TBP-ADN en amont de leur interaction [323-325]. Il a pour but de les stabiliser en compétitionnant avec Taf1 de TFIID [316, 326]. Pour ce qui est de TFIIB, il est composé d'une seule sous-unité [327, 328]. Il est en interaction directe avec l'ADN en amont du TSS via un motif appelé BRE (*B recognition element*) [329]. Ses extrémités N-ter et C-ter ont des fonctions différentes [330, 331]. Le N-ter est composé des domaines *B-ribbon*, *B-reader* et *B-linker*, et il est impliqué dans le recrutement de l'ARNPII et dans la formation de la bulle de transcription [332]. Ces trois domaines se localisent dans le site actif de l'ARNPII via le canal de sortie de l'ARN naissant [332-335]. Une fois à l'intérieur, le *B-linker* entre en interférence avec l'hybride ADN-ARN d'une longueur de 6 bp causant la séparation des deux molécules [332]. Puis, lorsque l'ARN atteint une longueur d'environ 12 bp, il y a interférence avec le *B-linker* et le *B-reader*, ce qui provoque la dissociation de TFIIB avec l'ARNPII [335, 336]. De plus, le *B-ribbon* est en interaction avec le *wall* et le *dock* de la polymérase et permet de localiser le métal B dans le site actif [332]. Pour ce qui est de la section C-terminal de TFIIB, elle est impliquée dans la stabilisation de l'interaction TBP-ADN [330, 331]. Contrairement à TFIID et à TFIIA qui ont une interaction stable avec l'ADN, TFIIB est en interaction transitoire et dynamique qui est, par la suite, stabilisée par le recrutement de TFIIF et de l'ARNPII [337].



**Figure 1.6 Schématisation de la formation du complexe de pré-initiation de la transcription.**

La reconnaissance de la boîte TATA au promoteur par TBP provoque un pli de 90 degrés dans l'ADN. L'interaction TBP-ADN est ensuite stabilisée par le recrutement de TFIIA et de TFIIB de part et d'autre du TBP. Le complexe TFIIF-ARNPII est ensuite recruté suivi de TFIIE et de TFIIH. L'arrivée d'ATP permet d'ouvrir la bulle de transcription par la translocase Ssl2 de TFIIH. Enfin, le premier lien phosphodiester est créé avec l'ajout de NTP. Figure adaptée de Sainsbury *et al.*, 2015 [319].

À la suite du recrutement de TFIIB, le complexe TFIIF-ARNPII est recruté. TFIIF est composé de trois sous-unités chez la levure (Tfg1, Tfg2 et Tfg3), tandis que celui de l'humain contient deux sous-unités (RAP74 et RAP30, homologue à Tfg1 et Tfg2, respectivement) [338, 339]. Son association avec l'ARNPII libre a pour but de prévenir une interaction non spécifique de la polymérase avec l'ADN [340]. Il joue aussi un rôle dans la formation du PIC en permettant une stabilisation de ce dernier avec TFIIB [336, 337, 341, 342]. De plus, il est aussi nécessaire à la formation de la bulle de transcription et au début de la synthèse de l'ARN [343, 344]. Il est en interaction avec l'ADN en amont du TSS.

L'avant dernier GTF à être recruté est TFIIE. Il est formé par les sous-unités TFIIE $\alpha$  et TFIIE $\beta$  et il est en interaction à la fois avec l'ADN, TFIIH et l'ARNPII [345-349]. Le rôle de ce facteur est principalement de permettre le recrutement de TFIIH et de stimuler son activité de phosphorylation [349, 350]. De plus, TFIIE $\beta$  s'accroche à l'ADN simple brin non transcrit afin de le stabiliser [348].

Enfin, TFIIH est recruté. Il est composé de dix sous-unités, dont trois ont des activités enzymatiques, c'est-à-dire une kinase (Kin28/CDK7), une hélicase (Rad3/XPD) et une translocase (Ssl2/XPB) [351]. TFIIH se localise dans le PIC entre le *hook* et le *shoulder* du Médiateur, ce qui permet de localiser la kinase près du CTD de l'ARNPII [352]. De plus, le module de la kinase est très mobile et il est stabilisé par le Médiateur, ce qui permet de stimuler la phosphorylation de la Ser5 du CTD [353-355]. L'hélicase Rad3 est principalement impliqué dans la réparation de l'ADN. Pour ce qui est de Ssl2, son activité ATP-dépendent permet l'ouverture de l'ADN à l'aide de TFIIE. Comme TFIID maintient l'ADN en aval de Ssl2, il y a une torsion qui permet d'ouvrir l'ADN [352, 354]. Après environ 10 bp de translocation, il y a une dissociation de l'enzyme [356]. Comme la bulle de transcription est très instable et que les deux brins d'ADN peuvent se réassocier, ils sont stabilisés par les facteurs TFIIB, TFIIE et TFIIF [354, 357, 358]. S'ensuit alors l'étape de balayage pour localiser le TSS. Il a été proposé que la bulle de transcription, lors de cette étape, soit de 6 bp et une fois le TSS identifié, la bulle s'agrandit à 13 bp chez la levure [359].

Le complexe que forme le PIC est stabilisé par le Médiateur, un méga complexe de 1,4 MDa [360]. Formé de 25 sous-unités chez la levure et de 30 chez l'humain, il est constitué de quatre modules : le *head*, le *middle*, le *tail* et la kinase. Les deux premiers sont impliqués dans l'interaction avec les GTF (TFIIB, TFIIE, TFIIF et TFIIH) et l'ARNPII [114, 361-364]. De plus, le Médiateur est aussi impliqué dans la régulation du recrutement des GTF TFIIA, TFIIB, TFIIF, TFIIE et TFIIH [365-368]. Mise à part sa contribution dans la formation du PIC, il joue aussi un rôle dans le relâchement de l'ARNPII en stimulant l'activité kinase de TFIIH. Le *tail* est quant à lui impliqué dans l'interaction avec les activateurs [360]. Cette liaison permet d'effectuer le pont entre les signaux de régulation et le PIC. En effet, le Médiateur permet un rapprochement des *enhancers* et du promoteur en formant une boucle dans l'ADN avec l'aide de la cohésine [369]. Chez la levure, les quatre modules du Médiateur se localisent à l'UAS via une interaction avec l'activateur [146, 147, 370]. Il a aussi été démontré récemment que la distribution du Médiateur à l'UAS est en partie possible grâce à TFIID. Une localisation transitoire au promoteur est aussi détectable, mais elle ne comprend pas le module de la kinase. En plus de son rôle majeur dans l'initiation, le Médiateur est aussi impliqué dans l'élongation, la maturation et l'exportation des ARN [371].

En somme, l'étape de l'initiation de la transcription débute lors de l'apparition du premier lien phosphodiester de l'ARN. Après une synthèse de 30 bp, l'ARNPII relâche le promoteur et les GTF pour entrer dans l'étape de l'élongation [20].

## 1.4.2 Élongation

L'étape qui suit celle de l'initiation est l'élongation. Cette dernière correspond à la phase de synthèse de l'ARN durant laquelle différents événements de maturation de l'ARN ont lieu de façon co-transcriptionnelle: l'ajout de la coiffe en 5' de l'ARN et l'épissage. Comme l'initiation, l'étape de l'élongation implique une régulation contrôlée afin de permettre une bonne coordination entre la transcription, la maturation de l'ARN et le passage de l'ARNPII au travers de la chromatine. Durant l'élongation, la polymérase subit différents événements de pause et une modulation de sa vitesse, ce qui permet une fenêtre d'opportunité

pour la régulation. De plus, différents facteurs sont recrutés pour permettre une processivité optimale de la polymérase et le glissement ou l'éviction d'histones afin que la polymérase puisse passer la barrière qu'est la chromatine. Différentes enzymes sont aussi recrutées pour catalyser des modifications post-traductionnelle des histones.

#### **1.4.2.1 Pause au promoteur proximal**

Depuis plusieurs années, il est bien établi que, chez les métazoaires, l'ARNPII subit un évènement de pause proximale au promoteur à une grande majorité des gènes avec des intensités différentes [372]. Les premières évidences démontrant que la transcription pouvait aussi être contrôlée durant l'élongation ont été effectuées dans les années 1980 [373, 374]. En effet, ces études ont démontré que l'inactivation d'un gène peut avoir lieu durant l'étape de l'élongation, et ce, de façon indépendante à l'initiation. Quelques années plus tard, la première observation de la pause proximale au promoteur chez la drosophile a été effectuée par l'équipe de John Lis [375, 376]. Ils ont démontré que le gène répondant aux chocs thermiques *HSP70* comportait des ARNPII en attente d'être activés, situés à 20-50 bp en aval du TSS. Par la suite, le même type d'évènement a aussi été confirmé chez les mammifères aux gènes *MYC* et *FOS* [377, 378]. On a ensuite confirmé que cette pause était présente à l'échelle du génome pour environ 30 % des gènes chez la drosophile et les mammifères, contrairement à la levure qui n'en contient pas [152, 154, 160, 379, 380]. Les gènes ciblés sont souvent ceux impliqués dans une réponse à un signal, comme le stress, un dommage à l'ADN, le développement et la différenciation cellulaire [372]. Il est important de noter que la pause n'implique pas une inactivation du gène, car il a été démontré que moins d'un pourcent des gènes ayant cette pause sont inactifs [379, 380]. Cette pause a principalement deux fonctions qui offrent une fenêtre d'opportunité pour la régulation [381]. En effet, elle permet de garder la chromatine ouverte afin de faciliter la liaison des facteurs de transcription. Il existe donc une corrélation entre les NFR et la pause, mais les NFR n'impliquent pas toujours une pause comme le démontre le cas de la levure où des NFR sont présents sans que cette pause y soit [382]. De plus, elle permet aussi une activation rapide à la suite d'un signal. Toutefois, ce ne sont pas tous les gènes en pause qui s'activent rapidement [383]. Il a aussi été démontré qu'elle permet

de synchroniser l'ajout de la coiffe en 5' des ARN [384, 385]. Récemment, l'utilisation de la drogue triptolide, un inhibiteur de la translocase XPB de TFIID, a permis de mesurer le temps de rétention de l'ARNPII au site de pause proximale au promoteur [386-389]. Il a été déterminé qu'elle dure généralement quelques minutes, mais il est aussi possible d'observer des rétentions de quelques heures.

Les processus qui induisent cet événement sont moins bien compris. On sait toutefois que cela implique des facteurs négatifs à l'élongation. En effet, l'utilisation de la drogue DRB a permis d'identifier les facteurs clés, c'est-à-dire la kinase P-TEFb, et les facteurs *negative elongation factor* (NELF) et *DRB sensitivity-inducing factor* (DSIF) [157, 158, 390, 391]. Il a été démontré de façon *in vitro* que la pause peut être induite par NELF et DSIF après une synthèse de 30-40 nt de l'ARN [157]. Les deux facteurs sont recrutés précocement lors de l'élongation via, entre autres, une interaction avec l'ARN naissant [392, 393]. Ces trois facteurs sont nécessaires pour la survie et le bon développement des organismes supérieurs comme la drosophile, le poisson-zèbre et la souris [394-397]. Avec un mutant n'ayant pas DSIF ou NELF chez la drosophile ou les mammifères, il est possible d'observer de façon importante une diminution de la pause, mais elle n'est pas complète [154, 379, 398]. De plus, des études ont démontré que certains gènes sont plus sensibles à NELF que d'autres [152, 399].

P-TEFb est le premier facteur à avoir été identifié pour être sensible à l'inhibiteur de la transcription 5,6-Dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole (DRB) [390]. La kinase Cdk9 du complexe a comme substrat la Ser2 du CTD de l'ARNPII ainsi que les facteurs NELF et DSIF. La phosphorylation de NELF permet sa dissociation du complexe d'élongation et la phosphorylation du domaine CTR de Spt5 (DSIF) rend le complexe positif à l'élongation [166, 400]. Enfin, la phosphorylation de la Ser2 du CTD permet le recrutement de facteurs de l'élongation, comme PAF1C et Spt6 [65].

Le facteur NELF, composé de quatre sous-unités (NELF-A, -B, -C et -E), est important pour la pause proximale au promoteur [157, 160]. Présent chez les métazoaires, il est toutefois absent chez les nématodes, la levure et les plantes [160]. De récentes structures de

NELF avec l'ARNPII ont démontré que le facteur effectue une interaction avec le *funnel* de la polymérase, où se situe le pore d'entrée des NTP et le site catalytique [401]. Il y a aussi une interaction avec la boucle *trigger* de la polymérase qui inhibe la mobilité de cette dernière. NELF est aussi en interaction avec un domaine de l'ARNPII qui prévient le recrutement du facteur TFIIS. Cette interaction est très importante, car TFIIS est impliqué dans un processus qui permet de sortir la polymérase de son statut de pause [381]. Enfin, le facteur possède deux domaines agissant comme des tentacules qui sont en interaction avec DSIF et l'ARNPII. Il a été démontré que l'un de ces domaines est impliqué dans la pause et c'est celui-ci qui est phosphorylé par P-TEFb pour permettre la dissociation du facteur [160, 400]. Une fois éjecté du complexe, sa réassociation est prévenue par la liaison du complexe PAF1C. NELF requiert DSIF pour sa fonction qui est uniquement dans la région du promoteur, comme l'indique sa localisation génomique [157, 402]. De plus, le relâchement de NELF nécessite l'interaction de son RRM avec un ARN de l'*enhancer* (ARNe) [403]. Cette interaction induirait un relâchement de NELF avec l'ARN naissant permettant ainsi la dissociation du facteur.

Pour ce qui est de DSIF, il est composé de deux sous-unités : Spt4 et Spt5 [158]. Il se situe autour du *cleft* et du canal de sortie de l'ARN naissant de l'ARNPII [404]. Il est aussi en interaction avec le *wall* et le *clamp* de la polymérase. De plus, il comprend deux domaines qui agissent de pince pour se lier à l'ADN et à l'ARN [405]. Ces liaisons permettent de maintenir la bulle de transcription, de positionner l'ADN en aval et de retenir l'ARN dans le pore de sortie. De plus, le CTR de Spt5 est positionné très près du canal de sortie de l'ARN et du CTD de l'ARNPII, un endroit idéal pour le recrutement de facteur. Non nécessaire à la pause, le CTR est impliqué dans le recrutement de facteur, comme la machinerie de coiffage de l'ARN [406, 407]. Ainsi, il est proposé que l'évènement de pause proximale au promoteur agit comme une fenêtre temporelle pour permettre l'ajout de la coiffe à l'ARN de façon optimale [372]. Les données de structure démontrent que DSIF se localise au même endroit que les GTF TFIIB et TFIIE, ce qui permettrait d'empêcher leurs réassociations [253].

Dernièrement, on a aussi démontré que le complexe PAF1C est impliqué dans la pause proximale au promoteur. Premièrement, l'équipe de Robert Roeder a démontré que ce complexe est recruté par P-TEFb aux sites de pause [408]. Cet évènement permettrait ensuite

de stabiliser la kinase à cet endroit. De plus, la phosphorylation de DSIF par P-TEFb est aussi impliquée dans le recrutement de PAF1C [409-411]. Comme il l'a été décrit précédemment, ce complexe permet d'éviter une réassociation de NELF et il est aussi impliqué dans l'activation de DSIF. Deuxièmement, le groupe d'Ali Shilatifard a démontré qu'une perte de PAF1C induit une augmentation du relâchement de la pause [412]. Il a par la suite démontré que le complexe peut aussi se lier à l'*enhancer* et que ceci préviendrait une activation totale du gène et du relâchement de la pause par l'ARNe [413].

Gdown1 est aussi connu pour être impliqué dans cet événement. Il a été démontré qu'initialement ce facteur, souvent présenté comme la treizième sous-unité de l'ARNPII, réprime les activateurs de la transcription de façon *in vitro* et cela peut être renversé par le Médiateur [414]. Toutefois, il ne peut pas s'associer avec le PIC [415]. Plus récemment, il a été démontré que Gdown1 est impliqué dans une augmentation de la pause proximale au promoteur [416]. Ainsi, il se localise principalement au promoteur des gènes. De plus, il est en compétition avec TFIIF pour une interaction avec l'ARNPII [367, 416]. Des résultats de structure ont démontré que Gdown1 se localisait au même endroit que TFIIB et TFIIF sur l'ARNPII [417]. Comme il en sera discuté dans la prochaine section, TFIIF est aussi impliqué positivement dans l'élongation précoce [418-420]. Gdown1 servirait donc d'antagoniste à TFIIF pour inhiber l'élongation. Toutefois, le remplacement du TFIIF par Gdown1 est très lent et nécessiterait une pause déjà établie par NELF et DSIF [421].

#### **1.4.2.2 Facteurs d'élongation**

Durant l'élongation, un recrutement dynamique de divers facteurs a lieu. Ces derniers permettent d'activer l'élongation à la suite d'un stimulus, d'augmenter la vitesse d'élongation et de permettre une processivité optimale de l'ARNPII. Ces facteurs sont aussi importants pour redémarrer l'ARNPII après une pause et pour aider l'enzyme à passer la barrière qu'est la chromatine. La drogue 6-azauracile (6AU) a souvent été utilisée pour identifier ces facteurs d'élongation. Un traitement au 6AU provoque une diminution des GTP au niveau cellulaire, puis lorsque combiné avec la suppression d'un gène codant pour un facteur d'élongation, il y a



défaut de processivité et de croissance [164, 422, 423]. Cependant, il faut tenir compte que ce ne sont pas tous les mutants de ces facteurs qui possèdent une sensibilité au 6AU [55]. De plus, lors de la transcription, l'ARNPII fait face à un obstacle physique majeur, la chromatine. Cette dernière est composée de nucléosome autour duquel est enroulé l'ADN. Chaque nucléosome est constitué d'un tétramère (H3-H4)<sub>2</sub> et de deux dimères H2A-H2B formant un octamère [424]. Lors de la transcription, deux types de facteurs permettent ce passage : les remodeleurs de la chromatine et les chaperonnes d'histone [280, 425]. Les remodeleurs sont ATP-dépendants et ils permettent un glissement, une expulsion ou un changement de compositions des nucléosomes, tandis que les chaperonnes sont ATP-indépendants et ils permettent l'expulsion ainsi que la réassociation des nucléosomes pour maintenir l'intégrité de la chromatine. Dans cette section, seulement les principaux facteurs d'élongation seront décrits.

#### **1.4.2.2.1 DSIF**

Comme il a été mentionné dans la section précédente, DSIF est d'abord impliqué dans la pause de l'ARNPII chez les métazoaires. Puis, une phosphorylation de sa sous-unité Spt5 par P-TEFb permet d'activer son rôle positif de l'élongation. Le complexe est présent durant l'étape de l'élongation comme le démontre sa localisation, contrairement à NELF qui se situe seulement dans la région promotrice [402]. Toutefois, une question demeure : quel changement subit DSIF pour passer d'un facteur négatif à un facteur positif lors de sa phosphorylation? Récemment, Vos *et al.* ont observé, à l'aide de structures, que lorsqu'il y a interaction entre DSIF, PAF1C et Spt6 avec l'ARNPII, les deux pinces de DSIF changent de conformation, ce qui serait nécessaire pour son activité positif [253]. Basés sur ces résultats, ils ont proposé que ces changements seraient impliqués dans le rapprochement de l'ADN et pour faciliter le passage de l'ARN dans le canal de sortie. De plus, ces résultats confirment l'évidence que PAF1C est un co-activateur de DSIF [426].

#### 1.4.2.2.2 TFIIF

En plus de son rôle dans l'initiation, le GTF TFIIF est aussi important durant les premiers instants de l'élongation [427, 428]. En effet, il est impliqué dans le début de l'étape de l'élongation, mais il est ensuite dissocié de la polymérase [418]. Son interaction avec l'ARNPII est dynamique et est nécessaire pour la formation du PIC ainsi que pour stimuler l'élongation de l'ARNPII [418, 429]. De plus, cette interaction est en dualité avec celle de Gdown1 et de la polymérase [417, 421]. Comme Gdown1 est impliqué dans l'établissement de la pause et que TFIIF est impliqué dans l'élongation précoce, il est suggéré que TFIIF est nécessaire pour sortir l'ARNPII de la pause proximale au promoteur en évincer Gdown1 du complexe transcriptionnel [430]. TFIIF se localise à la fois au promoteur et en aval de celui-ci chez l'humain [431]. Toutefois chez la levure, on observe seulement une localisation au promoteur [277]. Cette différence chez les deux organismes peut être expliquée par le fait qu'il n'y a pas de pause proximale au promoteur chez la levure.

#### 1.4.2.2.3 TFIIS

Durant l'élongation, l'ARNPII subit plusieurs événements de pause [379, 432]. Ils sont associés avec un mouvement de recul de la polymérase sur l'ADN (*backtracking*) suivi d'un arrêt de la transcription [433-435]. Ces événements peuvent être occasionnés par une erreur de la transcription, la présence d'ADN endommagé ou pour répondre à un obstacle, comme un nucléosome ou une protéine liant l'ADN [432, 436-438]. Afin de sortir de cette impasse, l'ARNPII doit effectuer un clivage de l'ARN libre issu de son recul, car ce dernier se lie dans un site précis du pore qui emprisonne la boucle *trigger* de la polymérase empêchant ainsi l'élongation [439, 440]. Pour ce faire, le facteur TFIIS doit être recruté afin de stimuler cette activité de clivage et de résumer l'élongation [434, 436]. Lors de son recrutement, TFIIS se localise sur le *funnel* de la polymérase et un de ces domaines entre dans le pore pour avoir accès au site actif. Cela a pour effet de déplacer l'ARN libre de 9 nucléotides provenant du *backtracking* et de libérer la boucle *trigger* afin de stimuler l'activité endonucléase de la polymérase [439, 441]. TFIIS peut aussi sortir l'ARNPII de sa pause proximale au promoteur,

car la longueur minimale requise pour l'ARN est de 13 nucléotides, mais comme il a été décrit précédemment, NELF bloque son recrutement [442].

#### **1.4.2.2.4 PAF1C**

Initialement découvert chez la levure, PAF1C est composé de cinq sous-unités, Paf1, Ctr9, Cdc73, Leo1 et Rtf1 [381, 443]. Ce dernier n'est pas toujours en association avec le complexe et peut avoir des fonctions différentes et indépendantes des autres sous-unités [444]. Mise à part son rôle dans la pause proximale au promoteur, le complexe est aussi impliqué dans l'élongation, la modification post-traductionnelle d'histones, la maturation de l'ARNm et dans la terminaison transcriptionnelle [445]. Il est en interaction avec la Ser2-P et la Ser5-P du CTD de l'ARNPII et DSIF [65, 253]. PAF1C permet de stimuler l'élongation de concert avec DSIF et requiert la phosphorylation de Spt5 par la kinase P-TEFb [426, 446]. Des données de structures démontrent que PAF1C permet d'effectuer un réarrangement de la pince à ADN de DSIF afin de faciliter le rapprochement de l'ADN en amont [253]. Comme attendu, sa localisation génomique se situe dans l'ORF du gène [447]. Il est aussi impliqué dans le recrutement de P-TEFb et du super complexe d'élongation (SEC) [400]. De plus, il est requis dans le recrutement de CDK12 chez la drosophile, une kinase de la Ser2 du CTD [169, 408]. Enfin, il permet le recrutement du dimère Rad6 et Bre1 nécessaire à l'ubiquitination de H2BK123 chez la levure (K120 chez l'humain) [448].

#### **1.4.2.2.5 FACT**

FACT (*facilites chromatin transcription*) est l'une des chaperonnes les mieux caractérisées. Il est impliqué dans la transcription, la réplication et la réparation de l'ADN [449]. FACT est formé de deux sous-unités (Spt16 et Pob3) et de Nhp6, qui y est vaguement associé [450-452]. Chez l'humain, il est composé de Spt16 et de SSRP qui est en quelque sorte une fusion de Pob3 et de Nhp6 [453]. Il se localise dans l'ORF des gènes, mais son recrutement est moins bien compris [131]. Deux mécanismes ont été proposés pour FACT afin de permettre le passage de l'ARNPII au travers de la chromatine durant l'élongation. Le

premier est l'expulsion d'un dimère H2A-H2B et le réassemblage du nucléosome à la suite du passage de la polymérase afin de maintenir l'intégrité de la chromatine [453-456]. Le deuxième modèle est un affaiblissement de l'interaction histone-ADN via une compétition entre FACT et l'ADN pour une interaction avec l'histone [457]. Récemment, il a aussi été démontré que FACT prévient l'insertion dans l'ORF du variant d'histone H2A.Z, connu pour border les sites NRF des promoteurs [458]. Un mutant de FACT occasionne donc un assemblage de ce variant dans l'ORF des gènes et provoque ainsi l'apparition de transcription cryptique.

#### 1.4.2.2.6 Spt6

Spt6, une autre chaperonne d'histone très bien caractérisée, est aussi responsable du réassemblage des nucléosomes et du maintien de l'intégrité de la chromatine durant l'élongation [456, 459, 460]. Comme FACT, il se localise dans l'ORF des gènes [131]. Il prévient aussi l'incorporation du variant d'histone H2A.Z et l'apparition de transcription cryptique [458, 461]. En plus de son rôle de chaperonne, Spt6 est aussi requis pour stimuler l'élongation [253, 462, 463]. Les récentes données de structure d'un complexe d'élongation (ARNPII, DSIF, PAF1C et Spt6) chez l'humain permettent de mieux comprendre ce rôle [253]. Comme attendu, Spt6 a besoin de PAF1C pour faciliter l'élongation et *vice versa*. Un complexe d'élongation sans DSIF est actif et peut transcrire, même si l'activité est plutôt faible. Toutefois, en l'absence de Spt6 ou de PAF1C, le complexe n'est pas actif. Il existe donc une certaine synergie entre les deux facteurs. Spt6 est en interaction avec Rpb4/Rpb7 ainsi qu'avec Spt5 (DSIF) et Ctr9 (PAF1C). L'impact de cette chaperonne sur le complexe d'élongation est un repositionnement des deux sous-unités de l'ARNPII et une ouverture de la pince à ARN de DSIF. De plus, Spt6 possède un domaine SH2 impliqué dans une interaction avec le *linker* du CTD de l'ARNPII ainsi que de ses résidus Ser2-P et Tyr1-P [251, 253, 254].

### 1.4.2.3 Évènements co-transcriptionnelles

Tout au long de la transcription, différents évènements reliés à la maturation de l'ARN sont effectués. Initialement, il avait été proposé que ces processus avaient lieu après la transcription, mais des évidences ont prouvé qu'ils étaient plutôt co-transcriptionnels [464]. Pour ce faire, les différentes protéines impliquées doivent être recrutées durant la transcription et cela est rendu possible par le CTD de l'ARNPII. En effet, le domaine agit comme plateforme pour le recrutement des facteurs liés à la maturation de l'ARN via ses différentes phosphorylations dynamiques et modulaires. Dans la prochaine section, il sera donc question de ces évènements : l'ajout de la coiffe aux ARN et l'épissage. La poly-adélynation et l'exportation seront abordés dans la section portant sur la terminaison transcriptionnelle.

#### 1.4.2.3.1 Coiffe des ARN

L'évènement de l'ajout d'une coiffe à l'ARN sert principalement à protéger ce dernier contre la dégradation par des exonucléases, comme Rat1 (XRN2 chez l'humain) [465]. Il y a aussi des évidences comme quoi cette coiffe est nécessaire aux autres processus de maturation de l'ARN comme l'épissage et l'exportation du transcrit dans le cytoplasme [466, 467]. De plus, elle est impliquée dans la traduction où elle serait, entre autres, un signal pour diriger la direction de la traduction [468].

L'ajout de la coiffe est le premier processus de maturation qui a lieu durant la transcription d'un gène. Il consiste à l'ajout d'un 7-méthylguanosine (m7G) à l'extrémité 5' de l'ARN naissant d'une longueur de 20-30 nt [469, 470]. La machinerie permettant cette réaction comprend trois activités enzymatiques [225]. La première correspond à l'hydrolyse d'un phosphate de l'ARN à l'aide d'une triphosphatase (Cet1). La deuxième étape consiste au transfert d'un GMP au 5' diphosphate de l'ARN par une guanylyl-transférase (Ceg1). Enfin, une guanine-7-méthyltransférase (Abd1) permet la méthylation de la guanine pour l'obtention finale d'un m7G-ARN. Le recrutement de la machinerie durant le tout début de l'élongation est rendu possible par la phosphorylation de la Ser5 phosphorylée du CTD de L'ARNPII, et ce, tout organisme confondu [226, 227, 230]. Toutefois, dépendamment des organismes,

d'autres modifications post-traductionnelles du CTD ou d'autres facteurs d'élongation peuvent aussi être nécessaires. En effet, chez les métazoaires, il a été démontré que la machinerie de coiffage requiert aussi le domaine en C-terminal du facteur d'élongation Spt5 du complexe DSIF pour son recrutement [471-473]. On propose aussi que la pause proximale au promoteur serve de point de contrôle au coiffage. Chez *S. pombe*, la mutation S5A du CTD est létale pour la cellule. Le groupe de Shuman a tenté de corriger cette létalité en fusionnant ce mutant du CTD avec l'enzyme de coiffage Mce1 (équivalent de Ceg1 chez *S. cerevisiae*), car il était déjà bien connu à l'époque que la Ser5-P du CTD est en interaction avec la guanylyl-transférase. Comme attendu, il a été possible de corriger la létalité du mutant S5A [82]. Toutefois chez *S. cerevisiae*, cette fusion du CTD muté avec Ceg1 ne permet pas de corriger la létalité [233]. Cela suggère que la Ser5 chez *S. pombe* est principalement requise pour le processus de coiffage ou d'autres événements non requis à la survie cellulaire, contrairement à *S. cerevisiae*.

#### 1.4.2.3.2 Épissage

Chez les eucaryotes supérieurs, les pré-ARNm sont constitués d'intron et d'exon. Comme seulement l'exon contient l'information nécessaire, l'intron doit être excisé et les différents exons doivent ensuite être raccordés. Afin de différencier ces segments, différents sites uniques se situent dans l'intron et l'exon. En effet, des sites d'épissage bordent les introns en 5' et en 3', et un site de branchement est aussi présent en 3' environ 20-40 nt en aval du site d'épissage en 3' [474]. De plus, l'exon contient aussi des séquences appelées amplificateurs d'épissage exonique qui permettent de recruter des protéines riches en sérines et en arginines (protéines SR). L'excision des introns nécessite deux réactions de transestérification qui sont rendues possibles par le spliceosome [475, 476]. Ce dernier est constitué de six snRNP et de plus de 150 protéines qui forment six complexes transitoires durant l'épissage [477, 478]. Avec les avancées technologiques de la cryo-microscopie électronique des dernières années, plusieurs structures de ces complexes ont été résolues chez l'humain et chez *S. cerevisiae*, ce qui permet une meilleure compréhension des mécanismes moléculaires [479-481].

Même si l'épissage peut être effectué indépendamment de la transcription, des imageries en microscopie électronique ont permis de confirmer qu'il a lieu de façon co-transcriptionnelle [482]. Le spliceosome est connu pour être en interaction avec le CTD et requiert aussi ce domaine pour effectuer l'épissage [483]. Les sous-unités U1 et U2 snRNP sont en interaction avec la Ser2-P et la Ser5-P du CTD via des domaines WW et FF [267, 484, 485]. De plus, le facteur d'épissage Prp40 contient un domaine WW qui est en interaction avec la Ser2-P et Ser5-P du CTD [267]. De plus, chez l'humain, les facteurs CA150, SF1, PSF, p54<sup>nrb</sup> et U2AF65 sont aussi connus pour avoir une interaction avec le CTD [268, 486-488]. Récemment, il a été démontré qu'une densité élevée d'ARNPII est localisée au site de jonction intron-exon et la Ser5-P du CTD est enrichie à la fin 3' de l'exon [489-491]. Cette marque serait nécessaire au recrutement du spliceosome [134]. De plus, il a également été suggéré que la vitesse d'élongation est importante dans la régulation de l'épissage alternatif. En effet, l'utilisation de mutants de l'ARNPII, connus pour être rapides ou lents, occasionne l'exclusion ou l'inclusion d'exon, respectivement [492, 493].

## 1.5 Transcription: Terminaison

Contrairement à l'initiation et à l'élongation, la terminaison transcriptionnelle est l'étape la moins bien comprise, et ce, même si des modèles ont été proposés, car beaucoup de détails mécanistiques sont manquants. La définition de cette étape consiste au relâchement de l'ARN transcrit et de l'ARNPII. Pour ce faire, les facteurs d'élongation doivent être relâchés, des facteurs de terminaison doivent être recrutés et la déphosphorylation du CTD doit avoir lieu pour permettre le recyclage de la polymérase. Une mauvaise terminaison peut se résumer à une terminaison précoce, à une accumulation d'ARN non fonctionnel et à l'envahissement du gène voisin [461, 494]. Contrairement à l'initiation qui se situe à un endroit précis du promoteur, la terminaison ne possède pas de position précise [495]. Chez *S. cerevisiae*, les sites de terminaison (TTS) se localisent en général à 150 bp du site poly(A), tandis que chez les métazoaires, ils peuvent être entre 100 bp et 1 kb du site poly(A) [496-498]. Dès la fin des années 80, deux groupes distincts ont émis l'hypothèse que la terminaison de l'ARNPII aux gènes codants s'effectuait via deux mécanismes [499, 500]. Le premier consiste au

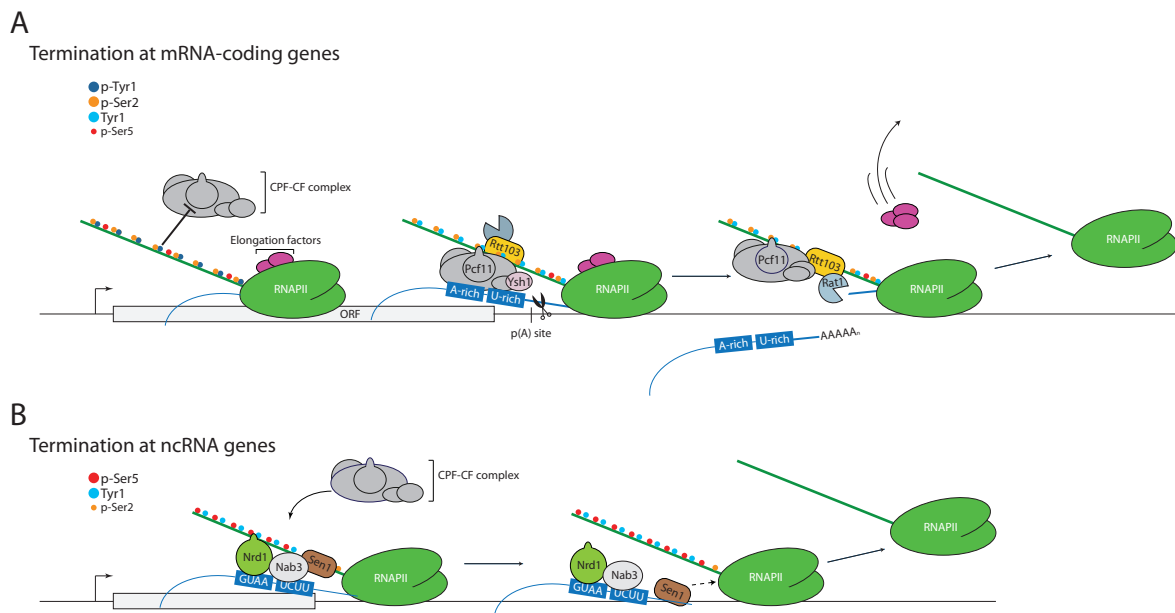
relâchement des facteurs d'élongation à l'aide d'un changement de conformation du complexe d'élongation lors de son passage sur le site poly(A). Le deuxième propose que l'ARNPII relâche l'ADN à la suite de la dégradation de l'ARN naissant par une exonucléase. Ces modèles ont, par la suite, été étudiés davantage et ils ont été confirmés. Ils portent d'ailleurs le nom de modèle allostérique et de modèle *torpedo*, respectivement [495, 501-503]. Puis avec les années, différents groupes ont découvert qu'il existait d'autres mécanismes de terminaison selon l'identité du gène. En effet, les gènes codants possèdent un mécanisme unique qui est très bien conservé dans l'évolution, où la maturation du pré-ARNm est intimement liée à la terminaison [504-506]. Puis, les gènes courts non-codants ont quant à eux un autre mécanisme, impliquant d'autres protéines, qui ne sont pas conservé entre *S. cerevisiae* et les mammifères.

### 1.5.1 Terminaison transcriptionnelle chez *S. cerevisiae*

Chez *S. cerevisiae*, deux mécanismes de terminaison ont été démontrés (Figure 1.7). Le premier est pour les gènes codants et nécessite le complexe CPF (*cleavage and polyadenylation factors*), aussi appelé CPF-CF. Ce dernier, comme son nom l'indique, est impliqué dans le clivage et la poly-adélation du pré-ARNm. Il comprend plus de 14 protéines qui forment un méga complexe d'environ 1 MDa divisé en trois modules [507]. Le premier module, appelé polymérase, est constitué de cinq protéines, dont Pap1, la polymérase responsable de l'ajout de la queue poly(A) aux transcrits. Il contient aussi Cft1 qui forme le cœur du module et du CPF-CF. Un complexe, appelé *cleavage factor I* (CFI), est aussi associé à ce module et joue un rôle clé dans la sélection du site poly(A) [508]. Il est composé de Rna14, Rna15, Pcf11 et Clp1, où Rna15 est impliqué dans l'interaction avec l'ARN naissant [509]. Rna14 est requis pour l'association de Rna15 avec le reste du complexe. Pcf11 est, quant à lui, important pour une interaction du CPF-CF avec l'ARNPII via le CTD et Rpb2 [125, 128, 510]. Puis, Clp1 permet la liaison du CFI avec le module polymérase [511]. Le second module, appelé nucléase, contient trois protéines, dont l'endonucléase Ysh1 impliquée dans le clivage du pré-ARNm. Ce complexe est associé avec les deux autres modules à l'aide de Ysh1 et de Cft2 [507, 512]. Enfin, le troisième module, nommé phosphatase, comprend six



protéines, dont les deux phosphatases du CTD, Ssu72 et Glc7, et il est rattaché à Ysh1 du module nucléase par sa sous-unité Pta1. Ce module peut exister sous deux formes [513]. L'une d'elle est de six protéines et est associée avec le CPF-CF, tandis que l'autre, nommé APT (*associated with Pta1*) contient sept protéines et n'est pas associée avec le complexe. La différence entre les deux est la présence de Syc1. Des évidences démontrent que l'APT est plutôt impliqué dans la régulation transcriptionnelle des snoRNA. En somme, toutes ces protéines du CPF-CF se localisent principalement dans la région 3' des gènes codants et le complexe est recruté via des interactions avec l'ARN naissant et l'ARNPII [514-516].



**Figure 1.7 Schématisation des mécanismes de terminaison transcriptionnelle chez *S. cerevisiae*.**

A) Mécanisme de terminaison pour les gènes codants pour des protéines. Lors de l'élongation, la présence de la Tyr1-P du CTD de l'ARNPII inhibe le recrutement précoce du complexe de terminaison CPF-CF. Une fois dans la région du site poly(A), la Tyr1-P est déphosphorylée par la phosphatase Glc7 et en concomitance avec la Ser2-P, il y a recrutement de Rtt103 et du CPF-CF par une interaction entre Pcf11 et le CTD. Ce recrutement permet ensuite un clivage du pré-ARNm par l'endonucléase Ysh1 et la poly-adénylation du transcrit par Pap1, ce qui mène ensuite à deux évènements qui vont permettre la terminaison. Soit le relâchement des

facteurs d'élongation (modèle allostérique) et la dégradation de l'ARN naissant par l'exonucléase Rat1 (modèle *torpedo*). B) Mécanisme de terminaison aux gènes courts non-codants, comme les snoRNA et les CUT, ainsi que certains gènes codants dits d'atténuation. Ce modèle implique le complexe NNS (Nrd1-Nab3-Sen1). Nrd1 est recruté par la Ser5-P du CTD et par une interaction avec l'ARN, ce qui permet par le fait même le recrutement de Nab3 et Sen1. Ensuite, Sen1, via son activité hélicase, rattrape l'ARNPII pour induire la terminaison. Des évidences démontrent aussi l'implication du complexe CPF-CF, toutefois son rôle n'est pas encore connu.

Pour ce qui de l'implication de la maturation du pré-ARNm dans la terminaison, les premières études ont démontré que le site poly(A) y est nécessaire, sans quoi l'évènement n'a pas lieu [499, 500, 517]. Afin d'expliquer comment la terminaison a lieu, il a été proposé que l'ARNPII doit subir un changement de conformation lors du recrutement du complexe CPF-CF, appelé modèle allostérique [499]. Ce mécanisme a ensuite été confirmé de plusieurs manières. Premièrement, il a été démontré de façon *in vitro* que le recrutement de Pcf11 au complexe d'élongation provoque un relâchement des facteurs d'élongation [518]. Une autre étude a démontré que le facteur d'élongation PAF1C est relâché du complexe d'élongation au site poly(A) en amont du site de terminaison [514, 515]. Enfin, des évidences montrent que les sous-unités Rpb3 et Rpb11 de l'ARNPII sont aussi importantes pour la terminaison [519]. Pour expliquer cette dernière observation, les auteurs proposent que, lors de cette étape, un changement de conformation de l'ARNPII a lieu via ces deux sous-unités pour ensuite induire le relâchement de la polymérase. Pour ce qui est du deuxième modèle, il a lieu à la suite du recrutement du complexe CPF-CF et du clivage du pré-ARNm. En effet, lors du clivage par Ysh1, l'ARN naissant, encore accroché à l'ARNPII, est dégradé par une exonucléase, plus spécifiquement Rat1, un facteur se localisant dans la région du site poly(A) [234]. Ce mécanisme, appelé *torpedo*, consiste au rattrapage de l'ARNPII par Rat1 afin d'induire la terminaison. Un mutant de cette exonucléase n'a aucun impact sur l'activité de clivage, mais ce dernier doit avoir lieu pour permettre à l'enzyme d'effectuer sa fonction [234, 248]. Il est toutefois important de noter que ce n'est pas seulement la digestion de l'ARN naissant par une exonucléase qui est nécessaire. En effet, un mutant Rat1 peut être compensé par une autre exonucléase, appelé Xrn1, et même si l'ARN est dégradé, il n'y a pas de terminaison [520,

521]. Des évidences *in vitro* démontrent aussi que Rat1 à lui seul ne peut pas effectuer le relâchement de l'ARNPII, d'autres facteurs, encore inconnus, sont aussi nécessaires [522, 523]. De plus, certaines études proposent que les deux modèles, allostérique et *torpedo*, sont complémentaires pour la terminaison [520, 522]. Il ne fait donc aucun doute que des mutations ou une diminution dans les différentes composantes clés du CPF-CF (Pcf11, Rna15, Glc7, Ssu72, Ysh1) et de Rat1 causent des défauts majeurs dans la terminaison des gènes codants [142, 504, 516, 521, 524-526]. De plus, il est proposé, surtout chez les métazoaires, qu'un évènement de pause est requis en 3' des gènes pour induire la terminaison, afin de permettre à XRN2 (homologue de Rat1) de rattraper l'ARNPII [489, 490, 495, 527]. Des pauses similaires à certains gènes sont aussi visibles chez la levure, mais leurs rôles sont plutôt controversés [134, 432].

Pour ce qui est du deuxième mécanisme, il est beaucoup moins bien caractérisé que le précédent. Il implique principalement trois protéines qui forment un complexe : Nrd1, Nab3 et Sen1 (NNS). Celles-ci sont impliquées dans la terminaison des gènes non-codants, comme les snoRNA et les CUT (*cryptic unstable transcript*), mais ciblent aussi certains gènes codants agissant comme atténuateur de la transcription [236, 528-530]. Les snoRNA sont des ARN non-codants (ARNnc), très courts, nécessaires à la maturation des pré-ARN ribosomiaux [531]. Les CUT sont, quant à eux, des ARNnc très répandus qui chevauchent les gènes fonctionnels voisins [528, 532, 533]. Ils n'ont aucun rôle fonctionnel connu, mais leur terminaison précoce ainsi que leur dégradation rapide par l'exosome est très importante pour éviter une interférence avec le gène voisin [534]. En plus du fait qu'ils sont très faiblement exprimés, il est très difficile de les détecter en *Northern blot* ou en ARN-seq. Pour ce faire, des mutants de l'exosome doivent être utilisés pour les stabiliser. De plus, il a été démontré que les promoteurs chez *S. cerevisiae* sont bidirectionnels et que la transcription de l'anti-sens mène à l'apparition de CUT [533, 535]. Le NNS agit donc comme un système de surveillance pour permettre la terminaison de ces CUT suivis d'une dégradation rapide par l'exosome [528]. Il existe aussi d'autres ARNnc, comme les SUT (*stable unannotated transcripts*) et les XUT (*Xrn1-dependent unstable*), mais ils sont principalement ciblés par le complexe CPF-CF [528, 536, 537]. Enfin, des gènes codants sont aussi atténués par une terminaison précoce via le NNS [528]. Finalement, il a été proposé que certains gènes codants terminent de façon Sen1

dépendante [538], mais une étude récente utilisant une méthode de haute résolution (PAR-CLIP) a plutôt constaté que ces gènes sont indépendants [524].

Pour ce qui est du mécanisme, contrairement aux gènes codants, aucune évidence ne démontre que les ARNnc doivent subir un clivage pour permettre leur terminaison. Chacune des protéines du NNS se localisent dans la région 5' des gènes codants et non codants [128, 521]. Toutefois, les motifs d'ARN associés à la liaison de Nrd1 se concentrent principalement aux ARNnc [496, 528]. Il a été proposé que ce facteur de terminaison soit le premier à être recruté sur l'ARNPII via une interaction avec la Ser5-P du CTD et la sous-unité Rpb2 [495]. Il est aussi en interaction directe avec l'ARN naissant via le motif GUAA/G [539, 540]. Nab3 possède, quant à lui, un domaine d'interaction avec Nrd1, l'ARN naissant (UCUUG) et Sen1 [238, 539-541]. D'ailleurs, il a été proposé que son rôle est de permettre le recrutement de Sen1. Ce dernier est une hélicase ATP-dépendante qui est très peu abondante dans la cellule [542]. En effet, on retrouve seulement ~125-500 molécules/cellule (contrairement à Nrd1 qui est de ~2 000-20 000 molécules/cellule) et lorsque le facteur est surexprimé, il est dégradé par le protéasome [543-545]. Selon le modèle de terminaison, le complexe NNS est recruté pour permettre à Sen1 de rattraper l'ARNPII afin d'induire la terminaison, comme pour Rat1 [495, 501, 503]. Toutefois, contrairement à ce dernier, Sen1 à lui seul, via son activité hélicase, peut induire la terminaison *in vitro* [522, 546]. D'autres études *in vitro* ont démontré que Sen1 est une enzyme très peu processive. Conséquemment, elle doit être recrutée très près de l'ARNPII qui doit s'arrêter ou ralentir pour permettre à l'enzyme de la rattraper [547]. Des études *in vivo* ont aussi démontré la nécessité de la diminution de la vitesse d'élongation pour permettre cette terminaison [53]. Ils ont aussi montré, à l'aide d'expériences en RT-qPCR, qu'un mutant rapide de l'ARNPII occasionne des défauts de terminaison aux snoRNA. Toutefois, l'utilisation du même mutant en ChIP-seq par ce même groupe ne permet pas d'observer cet effet [548]. Récemment, il a été démontré que les sites de terminaison des snoRNA comportent des protéines liées à l'ADN, appelées *roadblock*, ou une grande densité de nucléosome [549]. Les auteurs proposent que ces éléments sont impliqués dans la pause ou le ralentissement de la polymérase. Enfin, il existe aussi plusieurs évidences selon lesquels certains membres du complexe CPF-CF sont aussi impliqués dans ce mécanisme de terminaison, comme Ssu72, Glc7 et Pcf1, mais on ne comprend pas encore leurs fonctions

[202, 521, 550, 551]. Pour Glc7, il a été proposé que la phosphatase forme un complexe avec le NNS pour permettre l'activation de Sen1 par sa déphosphorylation, mais aucun mécanisme n'a été démontré à ce jour [541].

Il est donc clair que le complexe NNS est aussi recruté en 5' des gènes codants, mais la terminaison n'est pas induite. Donc, comment le choix est-il fait entre les deux mécanismes? La présence de motifs sur l'ARN naissant et la phosphorylation du CTD de l'ARNPII sont présentées comme étant les éléments régulateurs de ce choix. En effet, des motifs d'ARN sont reconnus spécifiques par le CPF-CF et par le NSS. Pour ce qui est du CTD, la Ser5-P serait responsable du recrutement du NNS en 5' des gènes courts non-codants et la Ser2-P, avec l'aide de la Tyr1, dans le recrutement du CPF-CF dans la région 3' des gènes codants.

Pour la terminaison des gènes codants, il y a deux modifications d'importance majeure du CTD selon le modèle, soit la déphosphorylation de la Tyr1 en concomitance avec la phosphorylation de la Ser2, pour le recrutement de deux facteurs, Pcf11 et Rtt103. Comme il a été précédemment décrit, Pcf11 fait partie du complexe CPF-CF. Pour ce qui est de Rtt103, il est en interaction avec le dimère Rat1-Rai1 [234]. Il a été proposé que Rtt103 permette le recrutement de Rat1 au complexe d'élongation. Toutefois, avec l'utilisation d'un mutant n'ayant pas le gène *RTT103* (*rtt103Δ*) ou en effectuant le *anchor-away* du facteur, nous n'observons aucun défaut (données non publiées). Pcf11 et Rtt103 possèdent chacun un domaine d'interaction avec le CTD (CID) [235]. Contrairement au CID de Rtt103, celui de Pcf11 ne possède aucune liaison directe avec la Ser2-P [98, 235]. Cette dernière permet cependant de stabiliser la conformation en coude  $\beta$  du CTD, nécessaire à l'interaction avec Pcf11. Récemment, il a aussi été démontré que la Thr4 du CTD joue aussi un rôle clé dans l'interaction avec Rtt103, comme la Ser2 [134, 246, 247]. En effet, lorsque le CTD est phosphorylé à la Ser2, la Thr4 effectue une liaison avec son groupement phosphate, ce qui stabilise la structure. Ensuite, en absence de la Ser2-P, la Thr4-P est impliquée dans l'interaction avec le facteur. La structure initiale du CID de Pcf11 avec le CTD a aussi permis de constater que la Tyr1, lorsque phosphorylée, est en encombrement stérique avec le facteur [98]. Puis, lorsque déphosphorylée, un pont hydrogène entre le groupement hydroxyle de la Tyr1 et le CID permet de stabiliser l'interaction. Comme ce résidu du CTD est phosphorylé

durant l'élongation, il a été proposé que cette marque prévient un recrutement précoce des facteurs de terminaison pouvant être causés par la présence de la Ser2-P [128]. Lorsque l'ARNPII franchit le site poly(A), la Tyr1-P est déphosphorylée par Glc7, puis avec la Ser2-P, il y a recrutement du CPF-CF via Pcf11 ainsi que de Rtt103 [142]. Il est toutefois important de noter que cela est basé strictement sur des analyses à l'échelle génomique. De plus, l'utilisation d'un mutant de la kinase principale de la Ser2 (*ctk1Δ*) a permis de confirmer l'implication du résidu dans le recrutement de Pcf11 et Rtt103 [234, 515]. Toutefois, d'autres études tendent plutôt à démontrer que la Ser2 ne joue pas un rôle aussi central que le laisse présager les études et le modèle précédents [129, 552]. En effet, les défauts de terminaison de l'ARNPII observés avec des mutants *ctk1Δ* ou S2A du CTD ne permettent pas d'observer de défauts majeurs. On observe des défauts d'environ 100 bp, contrairement à des mutants de Pcf11, de Ysh1 ou de Rat1 qui sont de plusieurs centaines de paires de base [515, 516, 552, 553]. En somme, tous ces articles confirment que le CTD joue certes un rôle important, mais que d'autres interactions sont aussi impliquées et importantes, comme celles avec Rpb2 et l'ARN naissant.

Pour ce qui est de Nrd1, il contient aussi un domaine CID, mais qui est, cette fois-ci, en interaction avec la Ser5-P et le groupement hydroxyle de la Tyr1 du CTD [245]. Une corrélation entre la distribution de Nrd1 et la Ser5-P est aussi observable en CHIP-chip et l'utilisation d'un mutant *nrd1ΔCID* provoque des défauts de terminaison aux snoRNA [128, 554]. Toutefois, l'importance de cette interaction est aussi controversée, car même si Tudek *et al* observent un défaut avec le mutant *nrd1ΔCID* en CHIP-chip, le groupe de Stephen Buratowski ne détecte aucun défaut avec ce même mutant en *Northern blot* [241]. Comme pour la Ser2-P, l'importance de la Ser5-P est peut-être donc surestimée, car Nrd1 est aussi en interaction avec Rpb2 et l'ARN naissant, ce qui peut compenser la perte d'interaction avec le CTD [496, 510, 539]. Enfin, des études proposent que la Ser2-P est aussi impliquée dans la terminaison par le complexe NNS, car elle permettrait le recrutement de l'hélicase Sen1 [550, 552, 555, 556].

Contrairement à *S. cerevisiae*, la levure *S. pombe* possède un seul mécanisme ciblant à la fois les gènes codants et les non codants : le complexe CPF-CF [137]. Nrd1 possède un homologue chez cette levure, appelé Seb1, mais il ne forme aucun complexe avec Nab3, Sen1 et Dbp8 (paralogue de Sen1) [557]. D'ailleurs, ces trois protéines ne sont pas impliquées dans la terminaison des gènes codants et snoRNA [137]. C'est plutôt Seb1 avec le CPF-CF qui y est requis [137, 558]. Comme Nrd1, Seb1 contient un CID, mais au lieu d'être en interaction avec la Ser5-P, il est plutôt spécifique à la Ser2-P [558]. De plus, il existe une corrélation entre la Ser2-P, la Tyr1-P, Pcf11, Seb1 et la terminaison. Donc, mise à part son homologie avec Nrd1, Seb1 est très différent. Contrairement à *S. cerevisiae*, le mécanisme de terminaison chez *S. pombe* aux snoRNA nécessite un clivage de l'ARN par Ysh1 et, ensuite, une poly-adénylation [558]. Enfin, un modèle *double-torpedo* a été proposé chez cet organisme basé sur l'observation de défaut de terminaison avec des mutants de l'exosome [559]. Selon celui-ci, lorsque l'ARNPII subit une pause pour ensuite effectuer un mouvement arrière (*backtracking*), le bout 3' de l'ARN rendu accessible est dégradé par l'exosome, ce qui induit la terminaison. Cette étape peut aussi être en concomitance avec la dégradation de l'autre bout en 5' de l'ARN par Rat1 afin d'induire la terminaison. Un tel mécanisme peut expliquer l'effet de l'exosome sur la terminaison de certains gènes chez *S. cerevisiae* [560].

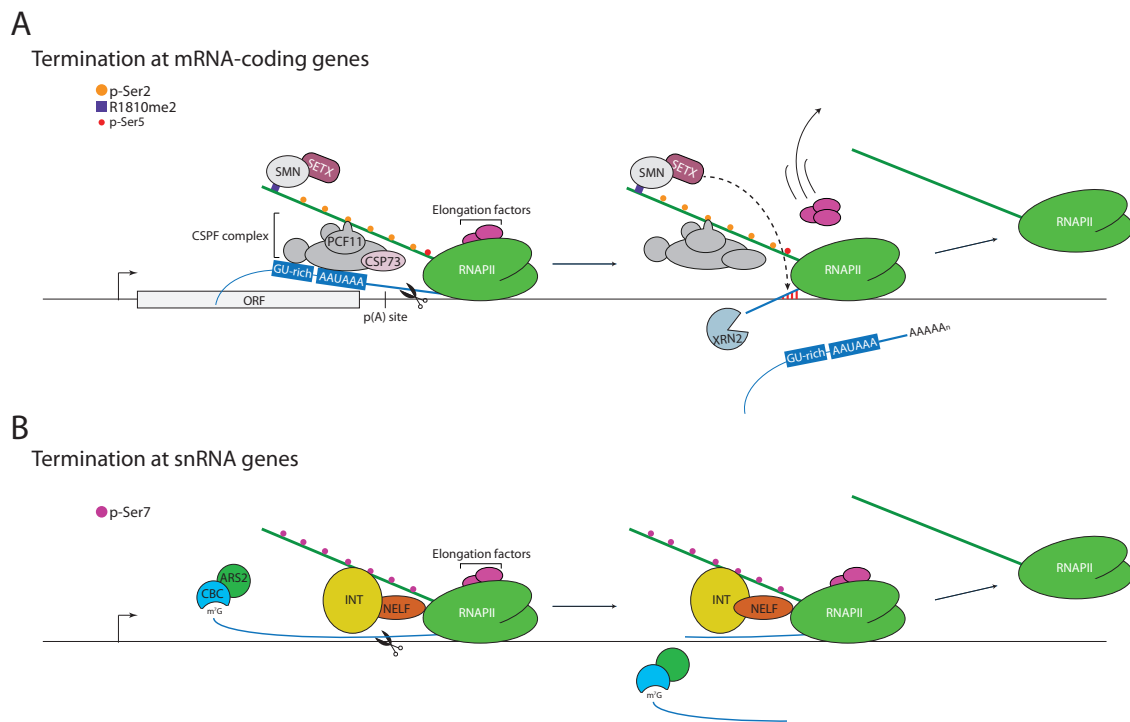
### 1.5.2 Terminaison transcriptionnelle chez les mammifères

Le mécanisme de terminaison pour les gènes codants chez les mammifères est très similaire à celui chez *S. cerevisiae* (Figure 1.8). En effet, il implique un complexe homologue au CPF-CF, appelé *cleavage and polyadenylation specificity factor* (CPSF) [503]. Ce dernier est aussi associé à d'autres complexes : le *cleavage stimulatory factor* (CstF) et un complexe formé de *cleavage factor I* (CFI) et de *cleavage factor II* (CFII). CFII est similaire au CFI de *S. cerevisiae*, où l'on retrouve Pcf11 dans les deux cas [561]. Donc, selon le modèle, le CPSF est recruté par l'ARNPII de façon CTD indépendante et via le signal poly(A) sur l'ARN, et le CFII via une interaction avec le CTD [562, 563]. Malgré que PCF11 est deux fois plus volumineux que son homologue chez la levure, il y a une grande homologie entre leur CID [564]. Ainsi, même si aucune étude ne l'a démontré, il est établi que le recrutement de PCF11

par le CTD est dépendent de la Ser2-P. De plus, il a été démontré que PCF11 peut effectuer le démantèlement du complexe d'élongation de façon *in vitro*, comme chez la levure, appuyant ainsi le modèle allostérique [565]. Pour ce qui est du complexe CstF, il est recruté par une interaction avec l'ARN naissant via un motif riche en GU ou en U [563]. Une fois tous les facteurs recrutés, le clivage du pré-ARNm est effectué entre 10 et 30 nt en aval du motif de poly(A) (AAUAAA) et en amont de la région riche en GU ou en U par CPSF73 (homologue de Ysh1) du complexe CPSF [566]. L'ARN naissant accroché sur l'ARNPII est ensuite dégradé par l'exonucléase XRN2 (homologue de Rat1) pour induire la terminaison (modèle *torpedo*) [567]. Il est intéressant de noter que des imageries électroniques ont permis de détecter des ARNPII, dans la région d'un site poly(A), liés à des ARN naissants courts, ce qui appuie le modèle qui veut que l'ARN doit être clivé avant que l'ARNPII soit relâché [568, 569]. De plus, des événements de pause sont aussi détectés dans la région du site poly(A) [490, 527, 568, 570]. Ces pauses permettraient à XRN2 de rattraper l'ARNPII. Ainsi, avec l'utilisation de mutants de l'ARNPII connus pour être lents ou rapides durant l'élongation, il est possible d'observer des terminaisons précoces ou tardives, respectivement [571]. Les pauses seraient, entre autres, causées par le recrutement séquentiel de la machinerie de clivage et de poly-adénylation sur la polymérase et l'ARN [562]. L'apparition de structure hybride d'ARN-ADN (*R-loop*) semble aussi jouer un rôle dans ces pauses [572, 573]. Il a d'ailleurs été démontré qu'elles sont ciblées par l'hélicase Senataxine (SETX), homologue de Sen1 [572]. Des mutations dans cette enzyme essentielle ont été caractérisées et démontrées pour occasionner des maladies graves, comme la sclérose latérale amyotrophique de type 4 (SLA4) et l'ataxie avec apraxie oculomotrice de type 2 (AOA2) [574-576]. Il a aussi été démontré que certaines des mutations menant à l'AOA2 occasionnent des défauts de terminaison aux gènes codants [576]. Enfin, le rôle proposé pour la SETX est que celle-ci est nécessaire pour catalyser la séparation des structures *R-loop*, ce qui ouvre le passage à XRN2 pour dégrader l'ARN naissant et induire la terminaison [572]. Chez la levure, il a été démontré de façon *in vitro* que Rat1, via son activité exonucléase, ne peut pas induire à lui seul la terminaison et qu'il nécessite un autre facteur encore inconnu à ce jour [522, 523]. Chez les mammifères, celui-ci semble être la SETX [572].



Récemment, un mécanisme de recrutement de la SETX impliquant le CTD de l'ARNPII a été proposé [215]. En effet, il a été démontré que la diméthylation de l'Arg1810 (Arg1810me2) du CTD par la méthyltransférase PRMT5 permet le recrutement du facteur SMN, via son domaine Tudor. Il est intéressant de noter que des mutations dans SMN sont impliquées dans la maladie d'amyotrophie spinale [577]. Les auteurs démontrent que l'interaction SMN-SETX est amplifiée par l'Arg1810me2 et est nécessaire pour leur recrutement sur l'ARNPII. Une diminution absolue d'une de ces trois composantes occasionne donc des défauts de terminaison et une augmentation de *R-loop*.



**Figure 1.8 Schématisation des mécanismes de terminaison transcriptionnelle chez les mammifères.**

A) Mécanisme de terminaison pour les gènes codants pour des protéines. Lors du passage du complexe d'élongation dans la région du site poly(A), le complexe de terminaison CSPF est recruté via une interaction avec le pré-ARNm naissant et avec la Ser2-P du CTD. Ce recrutement permet ensuite le clivage du pré-ARNm par l'endonucléase CSPF73 et la polyadénylation du transcrit par PAP, ce qui mène ensuite à deux évènements qui vont permettre le

relâchement de l'ARNPII, soit l'éviction des facteurs d'élongation (modèle allostérique) et la dégradation de l'ARN naissant par l'exonucléase XRN2 (modèle *torpedo*). Le recrutement du complexe CSPF permet aussi d'induire une pause à l'ARNPII, ce qui serait nécessaire afin de permettre à XRN2 de rattraper la polymérase. De plus, l'hélicase Senataxine (SETX) est recrutée par SMN à l'aide de l'Arg1810me2 du CTD de l'ARNPII, ce qui permet de résoudre les structures hybrides d'ARN-ADN, représentées en rouge. Cette étape serait requise afin de permettre à XRN2 de dégrader l'ARN naissant. B) Mécanisme de terminaison aux gènes snRNA. Dans ce mécanisme, la Ser7-P du CTD de l'ARNPII permet le recrutement du complexe intégrateur (INT). Ce dernier est aussi en interaction avec l'ARN. Ensuite, via ses endonucléases, INT clive le transcrit. Le facteur NELF est aussi impliqué dans le mécanisme par une interaction avec l'ARNPII et l'INT, mais son rôle exact n'est pas encore connu. Enfin, CBC-ARS2 est aussi important dans la terminaison via une interaction avec la coiffe de l'ARN.

Le deuxième mécanisme cible, quant à lui, les snRNA, les PROMPT (*promoter-proximal transcripts*, équivalents aux CUT chez la levure) et les gènes encodants des histones qui sont dépendants de la réplication (Figure 1.8) [578-581]. Les snoRNA ne sont toutefois pas inclus, car ils se situent pour la grande majorité dans des introns et ils ne possèdent pas leurs propres promoteurs et terminateurs, contrairement à chez la levure [582]. Ainsi, lors de la maturation du gène, l'intron excisé est ciblé par une endonucléase Rnt1 et par l'exosome pour permettre la maturation des snoRNA [583]. De plus, aucun homologue de Nrd1 et de Nab3 n'a été découvert jusqu'à présent chez les mammifères. Toutefois, Sen1 possède un homologue (SETX), mais il n'est pas impliqué dans la terminaison des gènes énumérés précédemment [576, 584]. Le mécanisme est beaucoup moins bien compris que celui nécessitant le CPSF. Il implique un complexe, appelé intégrateur, qui est en interaction avec la Ser7-P du CTD de l'ARNPII et un motif dans la région 3' de l'ARN [581, 585, 586]. L'intégrateur est constitué des endonucléases INT9 et INT11, homologue à CPSF73 et CPSF100 du CPSF [585]. Le complexe CBC-ARS2 est aussi impliqué dans cette terminaison via une interaction avec la coiffe de l'ARN naissant [580]. Il y a aussi des évidences que NELF est important via une interaction avec l'intégrateur et l'ARNPII [584, 587, 588]. Toutefois, aucun détail mécanistique n'a été décrit jusqu'à maintenant.

### 1.5.3 Évènements co-transcriptionnelles

#### 1.5.3.1 Clivage et poly-adénylation des ARNm

Tout comme l'initiation et l'élongation, l'étape de la terminaison comporte aussi des évènements co-transcriptionnels. En effet, comme il a été décrit dans la section 1.5.1 et 1.5.2, la maturation du pré-ARNm dans la région 3', c'est-à-dire le clivage et la poly-adénylation, est étroitement liée à la dernière étape de la transcription. Comme ces processus ont été décrits conjointement avec la terminaison dans les deux dernières sections, ils ne seront pas abordés dans celle-ci. Un autre évènement est, cependant, aussi lié à la terminaison. Il s'agit de l'exportation des ARNm dans le cytoplasme. Contrairement aux ARNnc, les ARNm doivent être transportés du noyau au cytoplasme afin d'être traduits. Pour ce faire, différents facteurs d'exportation sont nécessaires et certains sont recrutés de façon co-transcriptionnelle à l'aide du CTD de l'ARNPII [589, 590].

#### 1.5.3.2 Exportation des ARNm

Lors de la synthèse de l'ARN, différentes protéines sont associées aux pré-ARNm pour former des *messenger ribonucleoprotein* (mRNP) [589, 591]. Ceux-ci ont pour but de protéger les transcrits de la dégradation par l'exosome, de rendre l'ARNm apte à la traduction et surtout, de permettre le transport des transcrits dans le cytoplasme. Une fois que les mRNP sont complètement assemblés et que la terminaison transcriptionnelle a lieu, les mRNP sont dissociés de la machine transcriptionnelle et se déplacent vers le panier formé au pore du noyau où ils subissent un remodelage. Cela permet ensuite le relâchement des ARNm dans le cytoplasme ainsi que de recycler les facteurs de transport [592].

Un des facteurs principaux de l'exportation est la protéine Mex67 (NXF1 chez les mammifères) [593]. Cette dernière, qui forme un hétérodimère avec la protéine Mtr2 (NXT1 chez les mammifères), n'a pas de liaison directe avec l'ARN [594]. Pour ce faire, Mex67 requiert différents adaptateurs (Yra1, Nab2 et Npl3) qui sont recrutés lors de l'élongation

[595-598]. Ce sont d'ailleurs ces recrutements qui permettent de coupler l'étape de l'exportation avec la transcription. En effet, Yra1 est recruté à la machinerie transcriptionnelle par le CTD de l'ARNPII à l'aide de la Ser2-P et de la Ser5-P [269, 599]. Il a aussi été démontré que Pcf11 joue un rôle dans son recrutement lors de l'étape de l'élongation [600, 601]. Pour ce qui est de Nab2, il forme un complexe avec Yra1 et Mex67 et il est aussi en interaction avec la séquence poly(A) du transcrit [602]. Contrairement à Yra1 qui se dissocie du mRNP avant le transport dans le panier du pore, Nab2 est dissocié du mRNP dans le panier du côté du cytoplasme, comme Mex67 [603]. Enfin, Npl3 est une protéine faisant partie de la famille des protéines SR et elle est en interaction avec la Ser2-P du CTD de l'ARNPII [604]. Une déphosphorylation de Npl3 par Glc7 est toutefois nécessaire afin d'effectuer son transfert sur le mRNP [605].

Le complexe THO (Hpr1, Tho2, Thp2, Mft1 et Tex1) est aussi associé aux transports des messagers [589]. Il est recruté lors de l'élongation par une interaction directe avec la Ser2-P et la Ser5-P du CTD de l'ARNPII [590]. Une fois dans la région du site poly(A), l'adaptateur Yra1 se dissocie de Pcf11 pour se lier à l'hélicase Sub2 et, finalement, se connecter au complexe THO pour former le complexe TREX [598]. La liaison de ce complexe sur la chromatine requiert à la fois la Ser2-P et/ou la Tyr1-P du CTD de l'ARNPII [590]. De plus, lors de la poly-adénylation, la polymérase Pap1 du complexe CPF-CF est stimulée par les sous-unités Tho2 et Sub2, dans le cas contraire, le complexe TRAMP (Trf4-Air2-Mtr4) est recruté pour effectuer une poly-adénylation qui stimule ensuite la dégradation du transcrit par l'exosome [606]. Il est aussi important de noter que le recrutement de Mex67 sur la machinerie de transcription dépend de Hpr1 [607, 608]. En effet, ce dernier doit être ubiquitinylé par l'E3 ligase Rsp5 afin de permettre le recrutement de Mex67.

## **1.6 Problématiques, objectifs et méthodologies**

La plupart des mutants du domaine CTD de l'ARNPII sont létaux chez *S. cerevisiae*, c'est-à-dire les mutations Y1F, S2A et S5A, ce qui rend la tâche difficile pour étudier individuellement l'impact des résidus sur la transcription. Une des solutions pour être capable

d'utiliser ces mutants est l'utilisation d'expérience *in vitro*. Toutefois, même si ces études sont essentielles, elles ont des limites et des expériences *in vivo* doivent être effectuées. Pour ce faire, les mutants des kinases et des phosphatases du CTD étaient jusqu'à maintenant les principaux outils dans l'étude du rôle *in vivo* des différentes marques de phosphorylation du domaine lors de la transcription. Cependant, ces mutants ont plusieurs limitations, comme des défauts de croissance. Ils peuvent aussi posséder plus d'un substrat dont plusieurs résidus du CTD, ce qui est le cas pour la kinase Kin28 et la phosphatase Ssu72 [80, 130, 138]. En effet, Kin28 peut effectuer, entre autres, la phosphorylation de la Ser5 et la Ser7 du CTD. Il est donc difficile de partager le rôle de ces deux résidus avec l'utilisation d'un mutant de la kinase. La phosphatase Ssu72 cible aussi ces deux résidus du CTD. Un autre exemple est la phosphatase Glc7 qui possède plusieurs substrats dont la Tyr1 du CTD et l'hélicase Sen1 [541]. L'utilisation d'un mutant de cette enzyme pour étudier la terminaison est une limitation en soi, car il est impossible de savoir si l'effet se produit via le CTD ou Sen1. Il a été déterminé que Glc7 peut déphosphoryler Sen1 de façon *in vitro*, mais le rôle *in vivo* de cette action n'est toujours pas connu. De plus, certains résidus du CTD possèdent plusieurs kinases, comme la Ser2 qui est à la fois ciblée par Ctk1 et Bur1 [130, 172]. Ainsi, la perte d'une kinase peut être compensée par une autre. Une autre limite est que certaines de ces enzymes peuvent agir à titre d'imposteur lors de la perte d'autres kinases ou de phosphatases. En effet, il a été démontré que Bur1 (spécifique à la Ser2 du CTD) peut phosphoryler la Ser5 et la Ser7 lorsque Kin28 est absente [130].

L'objectif principal de mes travaux de doctorat est donc de caractériser chacun des résidus du CTD de l'ARNPII dans l'étape de la terminaison transcriptionnelle chez *S. cerevisiae* à l'aide de différents systèmes d'expression indépendants de l'utilisation des mutants de kinases et de phosphatases du CTD. Le deuxième objectif est de caractériser le rôle *in vivo* de la déphosphorylation de Sen1 par la phosphatase Glc7.

Pour être capables d'atteindre notre premier objectif, nous avons développé deux systèmes *in vivo* d'expression d'ectopique pour différents mutants du CTD, et ce, indépendamment de l'utilisation de ces mutants de kinases et de phosphatases. Pour ce faire, des mutants pour chacun des résidus phosphorylables du CTD ont été créés. La méthode

consiste à remplacer les résidus dans chacune des 26 répétitions par un acide aminé homologue ne comprenant pas le groupement hydroxyle, nécessaire à la phosphorylation. Par exemple, pour le mutant de la Tyr1, Y1F, chacune des Tyr1 a été remplacées par une phénylalanine. Il est important de noter que seulement les résidus du consensus ont été mutés, et que les vecteurs d'expression contiennent le promoteur et le terminateur endogène du gène *RPBI*. Les mutants ont ensuite été utilisés dans deux systèmes *in vivo*. Le premier consiste à l'expression ectopique de l'allèle *RPBI* contenant les différents mutants du CTD étiquetés FLAG dans des cellules exprimant de façon endogène le gène *RPBI* étiqueté MYC. La présence de l'endogène permet de garantir la survie cellulaire en présence de mutants létaux du CTD et comme les deux protéines sont différenciellement étiquetées, elles peuvent être analysées indépendamment. Le deuxième système, appelé CAND (*complementation after nuclear depletion*), utilise la même stratégie que le premier système sauf que, cette fois-ci, le gène *RPBI* contient une étiquette FRB au lieu de MYC afin de pouvoir utiliser une technique appelée *anchor-away* [146, 609]. Celle-ci consiste à sortir une protéine nucléaire du noyau à l'aide d'un traitement à la rapamycine. En effet, cette méthode est basée sur la dimérisation du domaine FRB de la protéine mTOR en C-terminal avec la protéine FKBP12 en présence de la drogue rapamycine. La stratégie consiste à étiqueter une protéine d'intérêt avec FRB et à fusionner FKBP12 à la protéine ribosomale RPL13A qui voyage dans le noyau pour sa maturation. Ainsi, avec la rapamycine, la protéine étiquette FRB se fusionne à RPL13A-FKBP12 et elle est expulsée du noyau après un certain temps (majoritairement 90 min.). Avant d'utiliser le système *anchor-away*, le gène exprimant l'homologue de mTOR (Tor1) doit être muté afin d'inhiber le potentiel toxique de la drogue. L'homologue de FKBP12 (Fpr1) doit aussi être muté pour éviter une compétition avec celui du système. En somme, ce deuxième système est principalement utilisé pour des expériences phénotypiques. Une fois les différentes souches obtenues, elles ont été utilisées dans des expériences d'immunoprécipitation de la chromatine hybridée à des micropuces (ChIP-chip) [610]. La technique *native elongating transcript sequencing* (NET-seq), qui consiste à immunoprécipiter l'ARNPII et à y extraire l'ARN naissant, a aussi été effectuée. Cette méthode est très puissante, car la résolution est très élevée (1 nucléotide) et elle ne cible que les polymérase actives. Elle permet aussi d'obtenir des résultats qui sont brins spécifiques. Enfin, des expériences des *Northern blot* ont aussi été effectuées pour valider les défauts de terminaison

observés afin de s'assurer qu'il ne s'agissait pas d'une induction de transcription cryptique par nos mutants. Pour ce qui est du deuxième objectif, la méthode *anchor-away* a été utilisée pour déterminer le rôle *in vivo* de Glc7 sur Sen1. Des expériences utilisant un système de dégradation inductible à l'auxine ont aussi été effectuées. Comme Sen1 est très peu exprimé, les analyses protéiques ont nécessité un enrichissement par précipitation afin d'éviter une surexpression.





**CHAPITRE 2: La Tyr1 du CTD de l'ARN polymérase II est requise pour une terminaison transcriptionnelle optimale par le complexe Nrd1-Nab3-Sen1**

## 2.1 Résumé

### **La Tyr1 du CTD de l'ARN polymérase II est requise pour une terminaison transcriptionnelle optimale par le complexe Nrd1-Nab3-Sen1**

Pierre Collin, Célia Jeronimo, Christian Poitras et François Robert

Chez *Saccharomyces cerevisiae*, la terminaison transcriptionnelle aux gènes codants est couplée au clivage du transcrit naissant, tandis que la terminaison de la plupart des gènes non-codants nécessite un mécanisme indépendant du clivage impliquant le complexe Nrd1-Nab3-Sen1 (NNS). Le recrutement des facteurs de terminaison implique des formes phosphorylées du domaine C-terminal (CTD) de l'ARN polymérase II, mais la contribution individuelle des résidus du CTD n'a jamais été examinée systématiquement. Pour ce faire, nous avons étudié l'impact des mutations des sites de phosphorylation du CTD dans la terminaison. Des défauts majeurs sont observés aux gènes codants avec les mutants de la Ser2 ou de la Thr4, tandis que des défauts de terminaison gènes spécifiques sont observés avec les mutants de la Tyr1. De plus, les mutations de Tyr1 provoquent des défauts de terminaison généralisés aux gènes non-codants se terminant par la voie NNS. Enfin, nous démontrons que la Tyr1 est importante pour la pause dans l'extrémité 5' des gènes et que le ralentissement de la transcription corrige les défauts de terminaison. Nos travaux mettent en évidence l'importance des pauses régulées par la Tyr1 dans la terminaison dépendante du NNS.

Article publié dans *Molecular Cell* (Facteur d'impact 2017 : 14,2)

## 2.2 Contribution

Conceptualisation : François Robert

Réalisation des expériences : Pierre Collin et Célia Jeronimo

Programmation informatique : Christian Poitras

Analyses des résultats : Pierre Collin

Écriture du manuscrit : Pierre Collin et François Robert

Supervision : François Robert

## 2.3 Manuscrit

### **RNA Polymerase II CTD Tyrosine 1 is Required for Efficient Termination by the Nrd1-Nab3-Sen1 Pathway**

Pierre Collin<sup>1</sup>, Célia Jeronimo<sup>1</sup>, Christian Poitras<sup>1</sup> and François Robert<sup>1,2,3\*</sup>

<sup>1</sup> Institut de recherches cliniques de Montréal, 110 Avenue des Pins Ouest, Montréal, Québec, Canada, H2W 1R7.

<sup>2</sup> Département de médecine, Faculté de médecine, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, Canada, H3T 1J4.

<sup>3</sup> Lead contact

Contact information:

\*Correspondence: [francois.robert@ircm.qc.ca](mailto:francois.robert@ircm.qc.ca)

### 2.3.1 Abstract

**In *Saccharomyces cerevisiae*, transcription termination at protein-coding genes is coupled to the cleavage of the nascent transcript, while most non-coding RNA transcription relies on a cleavage-independent termination pathway involving Nrd1, Nab3 and Sen1 (NNS). Termination involves RNA Polymerase II CTD phosphorylation, but a systematic analysis of the contribution of individual residues would improve our understanding of the role of the CTD in this process. Here, we investigated the impact of mutating phosphorylation sites in the CTD on termination. We observed widespread termination defects at protein-coding genes in mutants for Ser2 or Thr4 but rare defects in Tyr1 mutants for this class of genes. Instead, mutating Tyr1 led to widespread termination defects at non-coding genes terminating via NNS. Finally, we showed that Tyr1 is important for pausing in the 5' end of genes and that slowing down transcription suppresses termination defects. Our work highlights the importance of Tyr1-mediated pausing in NNS-dependent termination.**

**Keywords:** Transcription termination, RNAPII, CTD, Nrd1, Nab3, Sen1, Pcf11, Rtt103, Tyrosine 1, Pausing.

#### **Highlights**

- Tyr1 regulates termination of a small fraction of coding genes and of most non-coding transcripts
- Tyr1 mediates a pausing event in the 5' end of genes
- Tyr1-mediated pausing is required for efficient NNS-dependent termination
- Nrd1 and Sen1 mediate termination independently of their ability to bind the CTD

### 2.3.2 Introduction

Transcription termination is key in defining the boundaries of genes and has been found to play roles in transcript turnover and cell fate (Porrúa and Libri, 2015). Despite its importance, termination remains incompletely understood. In *Saccharomyces cerevisiae*, where transcription termination has been extensively studied, two mechanisms have been described, each operating at a different class of genes (Reviewed in (Kuehner et al., 2011; Mischo and Proudfoot, 2013; Porrúa and Libri, 2015; Richard and Manley, 2009)). At protein-coding genes, termination involves the cleavage of the pre-mRNA downstream of the polyadenylation (poly(A)) signal by the CPF-CF complex. The cleavage leads to the addition of the poly(A) tail to the cleaved transcript, but also triggers a cascade of events inducing termination within a window of a few hundreds base pairs downstream of the poly(A) site (Chanfreau et al., 1996; Jenny et al., 1996). A different termination mechanism has been described at short non-coding RNAs such as snoRNAs and cryptic unstable transcripts (CUTs). At these loci, termination involves the recruitment of Nrd1, Nab3 and the helicase Sen1 (NNS complex) (Arigo et al., 2006; Schulz et al., 2013; Steinmetz et al., 2001; Thiebaut et al., 2006; Vasiljeva and Buratowski, 2006). The helicase activity of Sen1 is thought to induce termination (Han et al., 2017; Hazelbaker et al., 2013; Porrúa and Libri, 2013), most likely by destabilising the RNA-DNA hybrid within the transcription bubble. Interestingly, the NNS pathway –unlike the poly(A)-dependent pathway– is not conserved in higher eukaryotes. Nevertheless, mammalian genomes code for a homolog of Sen1 (Senataxin) (Moreira et al., 2004). Interestingly, Senataxin is involved in poly(A)-dependent termination (Banerjee et al., 2009; Skourti-Stathaki et al., 2011; Suraweera et al., 2009; Wagschal et al., 2012), suggesting that evolution has incorporated some aspects of NNS-dependent termination in poly(A)-dependent termination.

In order to achieve efficient and accurate termination, the recruitment of elongation factors has to be tightly regulated in space and time. This is achieved via interactions of termination factors with specific sequences on the nascent RNA but also via interactions with the RNA polymerase II (RNAPII) C-terminal domain (CTD) (Porrúa and Libri, 2015). The CTD, which is a highly conserved and essential domain, is composed of 26 (*S. cerevisiae*) to

52 (mammals) repetitions of a heptapeptide having Tyr<sub>1</sub>Ser<sub>2</sub>Pro<sub>3</sub>Thr<sub>4</sub>Ser<sub>5</sub>Pro<sub>6</sub>Ser<sub>7</sub> as consensus (reviewed in (Corden, 2013; Eick and Geyer, 2013; Harlen and Churchman, 2017; Jeronimo et al., 2013; Jeronimo et al., 2016a; Zaborowska et al., 2016)). The tyrosine, serine and threonine residues of the CTD are dynamically phosphorylated during the transcription cycle, creating a very specific combinatorial phosphorylation pattern along genes (Bataille et al., 2012; Kim et al., 2010; Mayer et al., 2012; Mayer et al., 2010; Tietjen et al., 2010). This allows for the co-transcriptional recruitment of several factors involved in transcription and transcription-coupled processes (reviewed in (Corden, 2013; Eick and Geyer, 2013; Harlen and Churchman, 2017; Jeronimo et al., 2013; Jeronimo et al., 2016a; Zaborowska et al., 2016)). Studies have shown that the phosphorylation of Ser2 is required for the proper recruitment of Pcf11 (a component of the CPF-CF complex) and Rtt103 (a factor involved in recruiting Rat1) via their CTD-interaction domains (CID) (Ahn et al., 2004; Kim et al., 2004; Licatalosi et al., 2002; Lunde et al., 2010). More recently, *in vitro* studies have shown that Thr4 also contributes to the interaction of Rtt103 with the CTD (Harlen et al., 2016; Jasnovidova et al., 2017; Nemecek et al., 2017). Finally, X-ray crystallography and *in vitro* binding studies have shown that the interaction of Pcf11 and Rtt103 with P-Ser2 CTD peptides also involves a contact with the hydroxyl group of Tyr1 (Mayer et al., 2012; Meinhart and Cramer, 2004; Schrieck et al., 2014). Based on these data, it was proposed that the phosphorylation of Tyr1 would prevent premature recruitment of CPF-CF complex during elongation (Schrieck et al., 2014).

Evidence suggests that the NNS pathway relies on the phosphorylation of the CTD on Ser5. Indeed, a plethora of studies, including X-ray crystallography, have characterized an interaction between the Nrd1 CID and CTD peptides phosphorylated on Ser5 (Kubicek et al., 2012; Mayer et al., 2012; Tudek et al., 2014; Vasiljeva et al., 2008). Interestingly, the hydroxyl group on Tyr1 also makes important contacts with Nrd1 (Kubicek et al., 2012). These *in vitro* data are corroborated with ChIP experiments showing a correlation between the occupancy of Nrd1 and Ser5-P along genes (Mayer et al., 2012) and termination defects observed in the *nrd1ΔCID* mutant (Tudek et al., 2014). The importance of Ser5 in NNS-dependent termination is however controversial as no transcriptional readthrough was detected at the *SNR13* and *SNR33* loci by northern blot in the *nrd1ΔCID* mutant (Vasiljeva et al.,

2008). Finally, some studies also proposed a role for the CTD Ser2 in this pathway probably by recruiting Sen1 (Chinchilla et al., 2012; Grzechnik et al., 2015; Lenstra et al., 2013; Steinmetz et al., 2001).

While there is a lot of biochemical evidence for interactions between termination factors and the CTD (see above), understanding the functional role of these interactions in termination has been hampered by the fact that CTD mutations are often lethal or lead to very sick cells. Hence, the role of the CTD in termination *in vivo* has mostly been addressed using mutants of the various CTD kinases and phosphatases. These CTD-modifying enzyme mutants, however, are often sick. In addition, these enzymes often have multiple substrates or can act as imposters and compensate for each other (Akhtar et al., 2009; Bataille et al., 2012; Glover-Cutter et al., 2009; Kim et al., 2009; Nedeá et al., 2008; Qiu et al., 2009; Tietjen et al., 2010). In order to circumvent these limitations, and to address the role of the CTD more directly, we have developed a yeast strain where an ectopic *rpb1* allele —carrying the desired CTD mutation(s)— is expressed under the control of the *RPB1* promoter. This system allowed us to confirm the proposed role of Ser2 and Thr4 in termination of mRNA-coding genes in budding yeast. Surprisingly, we found that Tyr1 plays a much more limited role in poly(A)-dependent termination than previously suggested. Instead, Tyr1 appears to play an important role in termination via the NNS pathway, most likely by mediating a pausing event critical for termination.

### **2.3.3 Results**

#### **2.3.3.1 A dual-tag system to explore the role of the RNAPII CTD in transcription**

In order to study the role of CTD in transcription in conditions where CTD kinases and phosphatases are intact, an *in vivo* ectopic expression system was developed. This system consists of a Flag-tagged *rpb1* allele —carrying different CTD mutations— ectopically expressed in cells expressing the endogenous *RPB1*, which is Myc-tagged. The presence of the endogenous wild type (WT) *RPB1* allele allows maintaining cell viability and fitness, even in

the presence of ectopic alleles carrying otherwise lethal mutations. The endogenous and ectopic alleles are tagged with Myc and Flag respectively, allowing for their analysis by antibody-based assays such as western blotting, ChIP or NET-Seq. We introduced mutations in phosphorylatable residues in all 26 CTD repeats, turning them into non-phosphorylatable or sometimes phosphomimetic residues (see **Figure 2.1A**). For example, the Y1F mutant expresses an *rpb1* where all tyrosines are turned into phenylalanines. Note that all mutants were constructed within a WT CTD context, that is, in a natural repeat context (hence containing consensus and non-consensus repeats) rather than in an “all-consensus repeats” background. Importantly, all alleles were expressed at physiological levels and no sign of truncation or degradation was observed (**Figures S2.1A** and **S2.1B**). All mutations tested were viable in that context, although some, such as the Y1F mutant, exhibited slightly slower growth (**Figure S2.1C**). Chromatin immunoprecipitation (ChIP) experiments coupled to high-resolution tiling arrays (ChIP-chip) using anti-Myc and anti-Flag antibodies were used to monitor the genomic distribution of each mutant (Flag) together with its respective WT (Myc), both from the same chromatin extract. The genomic distribution of the different WTs (each expressed in the presence of a different mutant) were all very similar, highlighting no major dominant negative effect of the mutants on the WT enzyme (**Figure S2.1D**). Each mutant, however, showed noticeable differences when compared to their respective WT, indicating transcriptional defects. For example, at the 5’ end of genes, a notable difference was observed with the S5A mutant (**Figure 2.1B**), which has a profile consistent with the role of Ser5 phosphorylation in promoter escape (Jeronimo and Robert, 2014; Wong et al., 2014). In this report, we focused on the 3’ end of genes, investigating the role of the CTD in transcription termination.

### **2.3.3.2 Ser2 and Thr4 are globally required for efficient termination by the CPF-CF pathway**

At protein-coding genes, termination is coupled to cleavage and polyadenylation, and a plethora of work points to a key role of the phosphorylation of Ser2 in this process (reviewed in (Porrua and Libri, 2015)). As expected based on this prior work, our S2A mutant showed widespread termination defects. This is evident when looking at specific examples on a



genome browser (see pointing arrows in **Figure 2.2A**). In order to address the generality of the role of Ser2 (and other CTD residues) in termination at protein-coding genes, we computed a readthrough index (RTI) for all genes in all mutants. The index essentially measures the distance of the readthrough (see STAR Methods). As shown in **Figure 2.2B**, the RTI distribution is shifted to the right in the S2A mutant compared to WT (compare orange and black traces), indicating a global role for Ser2 in termination at protein-coding genes. Notice, however, that although general, the readthrough is rather small (less than 80 base pairs on average), indicating that termination still occurs in the S2A mutant but only with lower efficiency. This is consistent with termination occurring at more distal polyadenylation sites in the absence of Ser2 phosphorylation (Ahn et al., 2004; Skaar and Greenleaf, 2002). Interestingly, a similar phenotype is observed for the T4V mutant (light green trace in **Figure 2.2B** and examples in **Figure 2.2A**), highlighting a general role for Thr4 in termination at protein-coding genes. This is consistent with recent structural work showing that the hydroxyl group of Thr4 contributes to stabilising the interaction of Rtt103 with P-Ser2 peptides (Jasnovidova et al., 2017; Nemeč et al., 2017) and effects of T4V mutations on NET-seq profiles (Harlen et al., 2016). Note that the S5A and S7A mutants showed no sign of termination defects while Y1F showed a more complex RTI distribution (further described below).

We next tested whether the genome-wide termination defects observed in the S2A and T4V mutants were accompanied by decreased recruitment of termination factors, as predicted by prior work (Ahn et al., 2004; Harlen et al., 2016; Jasnovidova et al., 2017; Kim et al., 2004; Licatalosi et al., 2002; Lunde et al., 2010; Nemeč et al., 2017). For that, we used the complementation after nuclear depletion (CAND) system, where a mutant *rpb1* allele (here, carrying a CTD mutation) is expressed in cells where the endogenous *RPB1* gene is modified to be anchored-away from the nucleus upon rapamycin treatment (Jeronimo and Robert, 2014). This makes it possible to monitor phenotypic traits of any *rpb1* mutant following the nuclear depletion of the endogenous WT Rpb1 protein. As shown in **Figures 2.2C, S2.2A and S2.2B**, both the S2A and the T4V mutants showed decreased occupancy of both Pcf11 and Rtt103 after anchoring of WT Rpb1. Such a decreased in Pcf11 and Rtt103 occupancy was also observed in *ctk1Δ* cells and in a mutant where the S2A mutation was introduced in the

endogenous *RPB1* gene directly (Cassart et al., 2012) (**Figures S2.2C and S2.2D**). Taken together, our data on Ser2 and Thr4 confirm the bulk of literature on the role of these two CTD residues in termination at protein-coding genes and validate our system for the study of transcription termination.

### **2.3.3.3 Tyr1 is required for the efficient termination of a limited number of protein-coding genes**

We next looked at Y1F, a mutant where the hydroxyl group of Tyr1 is absent from all 26 repeats, hence preventing its phosphorylation but also abolishing any eventual contact between a CTD-binding protein and the hydroxyl group of Tyr1. Inspection of the ChIP-chip data revealed no apparent termination defects at most protein-coding genes in the Y1F mutant (see *PIR1* in **Figure 2.2A** and *MET10*, *CYC1* and *ADH1* in **Figure S2.2E** as examples), while a limited number of genes showed a major transcription readthrough (see *ZRT3* in **Figure 2.2.A** and *PIR3*, *TOM40* and *ZRG17* in **Figure S2.2F** as examples). Consistently with a gene-specific defect in termination at protein-coding genes, the RTI distribution of Y1F is multimodal, with a major peak centered at zero (indicating no defect) and a shoulder plus smaller peaks to the right (indicating termination defects) (**Figure 2.2B**, red trace). The gene-specific defect is also evident when looking at the data as a heat map of all genes (**Figure S2.3A**). In order to increase the resolution of our analyses of the Y1F mutant, we performed native elongating transcript sequencing (NET-seq) (Churchman and Weissman, 2011) and calculated a termination defect index (TDI) for each expressed gene (see STAR Methods). As shown in **Figures 2.2D**, **S2.2E** and **S2.2F**, the NET-seq experiment confirmed that only a limited number of protein-coding genes (67 genes out of 712 genes analyzed (9.4%); see STAR Methods and **Table S2.1**) show massive readthrough in the Y1F mutant. At genes suffering from termination defects in the Y1F mutant, transcription typically reads through several kilobases, invading the neighboring gene(s) (see **Figures 2.2A**, **2.2D** and **S2.2F**).

This Y1F termination phenotype was unexpected since phosphorylation of Tyr1 has been proposed to prevent early termination, limiting the recruitment of termination factors to a small window in the 3' end of genes (Mayer et al., 2012). According to this model, our Y1F

mutant should have showed pervasive defects, dominated by premature termination. A key evidence for this model is that nuclear depletion of the Tyr1 phosphatase, Glc7, led to RNAPII readthrough as observed in a metagene analysis of ChIP-chip experiments (Schrieck et al., 2014). Similar metagene analyses of i) our Y1F mutant (**Figure S2.3B**), ii) our own nuclear depletion of Glc7 (**Figure S2.3C**) or iii) the Glc7 anchor-away dataset from (Schrieck et al., 2014) (**Figure S2.3D**), all showed readthrough. RTI analysis of these Glc7 anchor-away datasets, however, shows bimodal distributions (**Figure S2.3E**), similarly to the Y1F mutant (**Figure 2.2B**), and the genes with readthrough in Glc7 depletion experiments are highly overlapping those with readthrough in the Y1F mutant (**Figure 2.2E**). Finally, replacing the 26 Tyr1 with either aspartic acids or glutamic acids –two phosphomimetics– also led to readthrough at a limited number of genes (see **Figure S2.3F** for examples), although the low level of RNAPII occupancy in these mutants prevented us from performing genome-wide analyses. Taken together, these experiments and analyses show that mutation of Tyr1 or nuclear depletion of the phosphatase Glc7 both lead to gene-specific transcription readthrough at a limited number of protein-coding genes and that the effect on these few genes dominated previous analyses (Schrieck et al., 2014).

#### **2.3.3.4 Tyr1 mutations lead to massive readthrough at non-coding RNA genes**

While inspecting our ChIP-chip and NET-seq data on a genome browser, we noticed that several non-coding RNA loci, known to terminate via the NNS pathway, showed massive readthrough in the Y1F mutant (see **Figure 2.3A** for examples). We therefore calculated RTIs and TDIs (using the ChIP-chip and the NET-Seq datasets respectively) for all annotated snoRNAs and cryptic unstable transcripts (CUTs) (excluding polycistronic and intragenic transcripts; see STAR Methods and **Tables S2.1** and **S2.2**). As shown in **Figures 2.3B** and **S2.4A**, massive readthrough dominates these non-coding genes in the Y1F mutant. These defects are specific to Tyr1 as they were not observed in other CTD mutants (**Figures 2.3A**, **S2.4B** and **S2.4C**). Northern blot analyses confirmed the presence of readthrough transcripts at all non-coding RNA tested in the Y1F mutant and also suggested that these transcripts are not processed properly (**Figure 2.3C**). Importantly, termination defects at snoRNAs in the Y1F

mutant highly overlap those found upon anchor-away of Nrd1 or Glc7 (**Figure 2.3D**). Interestingly, readthrough was also observed at protein-coding genes regulated by NNS-dependent attenuation (see *NRDI* in **Figure 2.3A**). Hence, we conclude that Tyr1 and its phosphatase are both generally involved in termination by the NNS pathway.

The fact that the depletion of Glc7 and mutating Tyr1 to phenylalanine (Y1F) both lead to very similar termination defects suggests that dephosphorylation of Tyr1 may be key for termination via the NNS pathway. Glc7, however, has other substrates, including Sen1 (Nedea et al., 2008), a key component of the NNS pathway (Nedea et al., 2008; Porrua and Libri, 2013; Vasiljeva and Buratowski, 2006). It therefore remains possible that Glc7 and the hydroxyl group of Tyr1 make independent contributions to NNS-dependent termination. In order to address this question, we tested two phosphomimetic mutants of Tyr1, namely Y1D and Y1E. These mutants mimic permanent phosphorylation of Tyr1 on all CTD repeats. Both Y1D and Y1E led to dramatic reduction in RNAPII occupancy over genes, suggesting a defect in early transcription steps (**Figure 2.4A**). This is perhaps due to impaired interaction between RNAPII and Mediator (Shah et al., 2018). Nevertheless, the Y1D and Y1E transcribing polymerases show very clear transcription readthrough at NNS-dependent genes as shown for a specific example in **Figure 2.4A**, as a metagene for snoRNAs in **Figure 2.4B** and by northern blot in **Figure 2.4C**. As an additional control, we introduced a catalysis-impairing mutation (D481A) (Werner and Weinzierl, 2002) in the Y1D or Y1E *rpb1* alleles, which led to the loss of the readthrough transcript from *SNR13* (**Figure 2.4C**), formally demonstrating that the phosphomimetic polymerases are transcription-competent and responsible for the transcription readthrough at NNS-dependent genes.

Tyr1 phosphorylation levels have been shown to increase with distance from the promoter at protein-coding genes in budding yeast (Mayer et al., 2012). Since most genes terminating via the NNS pathway are very short, we wondered whether P-Tyr1 is present at them. CHIP-chip experiments clearly showed that P-Tyr1 is present at these genes, despite their small size, and that it abruptly decreases at their 3' end, as shown in **Figure 2.4D**. This pattern is similar to what is observed at protein-coding genes (**Figure 2.4E**) (Mayer et al., 2012). Taken together, our data are consistent with dephosphorylation of Tyr1 by Glc7 being a

required step for termination by the NNS pathway although we cannot completely rule out other models (see Discussion).

### **2.3.3.5 NNS termination factor recruitment does not require their interaction with the CTD**

Phosphorylation of Ser5 has been implicated in NNS-dependent termination via the recruitment of Nrd1 (Gudipati et al., 2008; Kubicek et al., 2012; Tudek et al., 2014; Vasiljeva et al., 2008). As shown in **Figure 2.3A**, however, no obvious readthrough was observed in the S5A mutant (See also **Figures S2.4B** and **S2.4C**). We therefore tested the recruitment of Sen1 and Nrd1 using the CAND system. As shown in **Figures 2.5A**, **2.5B** and **S2.5**, specific CTD mutations (Y1F, S2A or S5A), or even the complete truncation of the CTD (CTDless), did not abolish the recruitment of NNS factors to snoRNA loci, despite leading to major transcription readthrough. Instead, Sen1 and Nrd1 appear to track with the readthrough polymerase in some mutants (**Figures 2.5A** and **2.5B**). Note that Sen1 and Nrd1 occupancy at snoRNAs was relatively reduced in the CTDless strain suggesting a contribution, although perhaps indirect, for the CTD in the recruitment of these factors. Interestingly, neither deleting the CTD-interacting domain (CID) of Nrd1 nor mutating a residue of Sen1 proposed to interact with the CTD (Chinchilla et al., 2012) had any measurable effect on the occupancy of Sen1 and Nrd1 at snoRNA loci (**Figures 2.5C** and **2.5D**). Strikingly, these mutations also did not lead to transcription termination defects at snoRNAs (**Figures 2.5C**, **2.5D**, **S2.5C** and **S2.5D**). Collectively, these data suggest that while complete occupancy of NNS components requires the CTD, their recruitment does not strictly require an interaction with the CTD. The data also suggest that the role of the CTD in this termination pathway goes beyond the mere recruitment of termination factors.

### **2.3.3.6 Tyr1 is required for RNAPII pausing in the 5' end of genes**

What is the mechanism through which the CTD, and more specifically Tyr1, mediates NNS-dependent termination? The data shown in **Figure 2.5** indicate that the recruitment of NNS factors is not abolished in the Y1F mutant. This suggests that the termination defect of

the Y1F mutant stems from failure of an event downstream of the recruitment of NNS factors. While inspecting the NET-seq data on a browser, we noticed that the peak of RNAPII typically observed towards the 5' end of genes, was greatly reduced in the Y1F dataset (See *PMA1* and *NRD1* in **Figures 2.6A** and **2.6B** for examples). Because, the excess of NET-seq signal in the 5' end of genes is generally attributed to pausing in that area (Churchman and Weissman, 2011), we further investigated whether Tyr1 may be involved in RNAPII pausing. First, we mapped the NET-seq data on the 5' and 3' ends of genes. This analysis clearly shows that the NET-seq signal is relatively depleted in the 5' end of genes in the Y1F mutant (**Figure 2.6C**). Heat maps of the NET-seq signal for all genes show that this effect of the Y1F mutant is global (occurring at all genes) but restricted to the 5' end of genes (**Figure S2.6A**). Despite being more frequent in the 5' end of genes, pausing occurs at various locations along genes (Churchman and Weissman, 2011). In order to investigate the contribution of Tyr1 on pausing more globally, we identified pause sites genome-wide, using the same algorithm developed by Churchman and Weissman (2011). This led to the identification of 23,045 pause sites (**Table S2.3**). Pause sites were then sorted based on their distance from the TSS and the NET-seq signal mapped around a 100 bp window around them. Interestingly, we found pause sites to be more dependent on Tyr1 as they get closer to the TSS (**Figures 2.6D** and **S2.6B**). This can also be observed at specific genes (see *GAS1* in **Figure 2.6E**). Taken together, these analyses show that Tyr1 is important for RNAPII pausing in the 5' end of genes, a feature that was recently observed in mammals (Shah et al., 2018).

Because NNS-dependent termination and attenuation is orchestrated early after initiation (snoRNA and CUTs are very short), we hypothesised that termination (and attenuation) by this pathway may take advantage of Tyr1-dependent pausing to achieve efficient termination. This would be in par with the fact that the Sen1 helicase is poorly processive (Han et al., 2017). Consistent with this possibility, Tyr1-dependent pause sites, located nearby Nrd1 binding sites, are readily visible at snoRNA and CUT loci (See **Figures 2.6F** and **2.6G** for examples). Hence, cells may leverage Tyr1-mediated pausing, early after initiation, to sustain efficient NNS-dependent termination of small non-coding RNAs or attenuation at some protein-coding genes.

### 2.3.3.7 Termination defects in the Tyr1 mutant are suppressed by slow RNAPII

A prediction of that model is that mutations or growth conditions that slow down transcription may suppress the termination defects of the Y1F mutant. We tested that model two ways. First, we treated Y1F cells with 6-azauracil (6AU), a nucleotide-depleting agent that imposes a stress on elongation (Exinger and Lacroute, 1992; Mason and Struhl, 2005). Second, we introduced mutations in the catalytic site of Rpb1 that are known to decrease synthesis rate (H1085Y and N488D) (Hazelbaker et al., 2013; Malagon et al., 2006; Malik et al., 2017). Quite strikingly, both the 6AU treatment and the *rpb1* mutations rescued the termination defect of the Y1F mutant at most NNS-terminated genes (**Figures 2.7A, 2.7B and S2.7A**) as well as at protein-coding genes where termination is Tyr1-dependent (**Figures S2.7B and S2.7C**). Interestingly, reducing the speed of RNAPII even suppressed the slow growth defect of our Y1F strain (**Figure 2.7C**). These data suggest that Tyr1 mediates a pausing event that is required for NNS termination. An alternative explanation, however, would be that the Y1F mutant is faster than WT, preventing NNS termination independently of pausing. Termination defects were indeed suggested to occur in fast RNAPII mutants (Hazelbaker et al., 2013). Unfortunately, the assay used to measure RNAPII velocity was recently shown to be erroneous when applied to faster than WT RNAPII (Malik et al., 2017), hence preventing us from directly measuring the speed of the Y1F mutant. We therefore addressed the question by looking at evidence for readthrough in fast RNAPII mutants. Neither our own ChIP-chip nor previously published ChIP-seq data (Soares et al., 2017) of fast RNAPII mutants showed readthrough approaching that observed in the Y1F mutant, however (**Figures S2.7D, S2.7E and S2.7F**). Hence, our data are consistent with Tyr1 and its phosphatase Glc7 regulating NNS termination by mediating pausing early after initiation (which is close to the site of termination due to the small size of snoRNAs and CUTs).

### 2.3.4 Discussion

The work presented here confirms the importance of the phosphorylation of Ser2 for the recruitment of the CPF-CF complex and Rtt103 to the 3' end of protein-coding genes. In addition, it highlights the contribution of Thr4 in that recruitment, as recently shown (Harlen

et al., 2016; Jasnovidova et al., 2017; Nemeč et al., 2017). Data gathered from our Y1F mutant –together with those from the anchor-away of Glc7–, however, call for a redefinition of the role of Tyr1 in poly(A)-dependent termination. Indeed, phosphorylation of Tyr1 was proposed to block binding of the CPF-CF complex, restricting its recruitment to the 3' end of genes, hence preventing premature termination (Mayer et al., 2012; Schreieck et al., 2014). A key experiment leading to this model was the observation that RNAPII reads through upon anchor-away of the Tyr1 phosphatase Glc7 (Schreieck et al., 2014). Re-inspection of those data, as well as our own, however, revealed that this readthrough only occurs at a small number of genes.

At snoRNAs and CUTs, we found that components of the NNS pathway are recruited independently of their interaction with the CTD. This finding was surprising since a plethora of *in vitro* data characterized an interaction between P-Ser5 CTD peptides and Nrd1 (Kubicek et al., 2012; Tudek et al., 2014; Vasiljeva et al., 2008). Nevertheless, however, we found termination to be CTD-dependent, suggesting that interactions between NNS factors and the CTD may be required for their function rather than their recruitment per se. The S5A mutant, however, did not show any detectable termination defect, contrarily to the currently accepted model (Kuehner et al., 2011; Porrua and Libri, 2015). This divergence is most likely explained by the difference in sensitivity between assays. Hence, while we do not rule out the possibility that Ser5 contributes to NNS-dependent termination, our work shows that its role is rather minor when compared to Tyr1 (see below).

The Y1F and the Glc7 anchor-away strains both show dramatic termination defects at NNS-dependent genes, leading to a large fraction of the RNAPII to readthrough. Based on visual inspection of our ChIP-chip and NET-seq data, these readthrough polymerases invade the neighboring gene(s) and eventually terminate at the next encountered poly(A) site. This leads to unprocessed snoRNA precursors, based on our Northern blot experiments, which probably explains the reduced fitness of the Y1F strain, despite the presence of an endogenous WT *RPB1* allele. What makes NNS-termination dependent on Tyr1? NNS factors recruitment does not strictly require their interaction with the CTD, suggesting that Tyr1 affects termination via a different mechanism. Since we observed a Tyr1-dependent pause at many



snoRNAs and CUTs and because Tyr1-dependent termination defects are suppressed by slowing down the polymerase, we propose that Tyr1 mediates termination by regulating a pausing event that is required for Sen1 to operate efficiently (**Figure 2.7D**). Tyr1-mediated pausing mainly occurs early after initiation (typically within 500bp of the TSS) but the small size of NNS-terminated transcripts positions these pause sites close to Nrd1 binding sites. This model is supported by *in vitro* work showing that Sen1 is a poorly processive helicase (Han et al., 2017).

Why do some protein-coding genes depend on Tyr1 for their termination? Interestingly, slowing down transcription rescued not only snoRNA and CUTs termination, but also termination at Tyr1-dependent protein-coding genes. This implies that the mechanisms through which Tyr1 operated is similar in both pathways. It therefore appears likely that the few protein-coding genes that –unlike most others– heavily rely on Tyr1 for termination do so because their termination requires a pause. It is worth noting that we were not able to identify any defining characteristic for the Tyr1-dependent protein-coding genes. They do not differ from other genes with regards to size, localisation along chromosomes, expression levels and chromatin structure (not shown). We also were not able to identify any particular sequence enriched in their 3' end that might have provided clues to their dependence on Tyr1 for termination. Interestingly, however, termination at these genes is Nrd1-independent, suggesting that the effect of Tyr1 on termination may go beyond NNS-dependent mechanisms. Hence, understanding why these genes require a Tyr1-mediated pause for efficient termination will require further investigation.

Recent work by Shah and colleagues (Shah et al., 2018) showed that in mammalian cells, Tyr1 is globally required for termination via the CPF-CF complex. Our work shows that in yeast, only a small fraction of protein-coding genes requires Tyr1 for efficient termination. These apparently opposing results may suggest a very different role for Tyr1 in yeast and mammals. On the other hand, the data are also consistent with more unifying scenarios. We privilege a model where the fundamental molecular function of Tyr1 is conserved but the consequence it has on termination differs. We propose that Tyr1 mediates pausing events in both organisms (This study and (Shah et al., 2018)). Such pausing events may then be more or

less critical for termination at different genes and organisms. In yeast, all NNS-dependent genes and some protein-coding genes would rely on pausing for efficient termination while in mammals, the requirement for pausing may be more generalized. This scenario would be consistent with the fact that mammalian cells depend on Senataxin for termination of protein-coding genes (Banerjee et al., 2009; Skourti-Stathaki et al., 2011; Suraweera et al., 2009; Wagschal et al., 2012). This scenario also better explains the observation that Tyr1 affects termination through different pathways (poly(A) and NNS); a fact that is more difficult to reconcile with models implying a role for Tyr1 in the recruitment of specific factors.

Phosphorylation of Tyr1 shows a very different pattern between yeast and mammals. In yeast, P-Tyr1 increases with distance from the TSS, reaching its maximum level near the 3' end of genes where it abruptly declines, just prior to P-Ser2 (Mayer et al., 2012). In mammals, however, P-Tyr1 is mainly found in the 5' end of genes (Descostes et al., 2014) making it more difficult to envision how it would regulate termination. This may argue for Tyr1 itself, rather than its phosphorylation, being important for pausing/termination. Another argument against the importance of phosphorylation stems from recent mass spectrometry analysis of yeast and human CTD that revealed very low level of Tyr1 phosphorylation in both organisms (Schuller et al., 2016; Suh et al., 2016). Yet, our data using phosphomimetic Tyr1 CTD mutants and Glc7 anchor-away strongly suggest –although do not prove– a role for CTD dephosphorylation in termination. Hence, while our work and that of others (Mayer et al., 2012; Schrieck et al., 2014; Shah et al., 2018) clearly established a role for Tyr1 in termination, whether or not this involves its phosphorylation/dephosphorylation will require further investigation.

### **2.3.5 Acknowledgments**

This work was funded by grants from the Canadian Institutes of Health Research (CIHR) to F.R. (MOP-133 648). This research was also enabled in part by support provided by « Calcul Québec » (<http://www.calculquebec.ca>) and Compute Canada ([www.computecanada.ca](http://www.computecanada.ca)). F.R. holds a Research Chair from the « Fonds de Recherche

Québec –Santé » (FRQS). P.C. held a studentship from the FRQS. We are grateful to N. Francis and F. Bachand for critical reading of the manuscript. We also thank L.S. Churchman and C. Trahan for advice on NET-seq library preparation, A. Almail for constructing plasmids pFR728 and pFR729, and D. Hermand, P. Cramer, D. Libri, R.A. Young and C.D. Kaplan for sharing yeast strains.

### 2.3.6 References

- 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.
- Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., and Ansari, A.Z. (2009). TFIIF kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* *34*, 387-393.
- 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.
- Banerjee, A., Sammarco, M.C., Ditch, S., Wang, J., and Grabczyk, E. (2009). A novel tandem reporter quantifies RNA polymerase II termination in mammalian cells. *PLoS One* *4*, e6193.
- Barnett, D.W., Garrison, E.K., Quinlan, A.R., Stromberg, M.P., and Marth, G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* *27*, 1691-1692.
- Bataille, A.R., Jeronimo, C., Jacques, P.E., Laramee, L., Fortin, M.E., Forest, A., Bergeron, M., Hanes, S.D., and 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.
- Brunelle, M., Coulombe, C., Poitras, C., Robert, M.A., Markovits, A.N., Robert, F., and Jacques, P.E. (2015). Aggregate and Heatmap Representations of Genome-Wide Localization Data Using VAP, a Versatile Aggregate Profiler. *Methods Mol Biol* *1334*, 273-298.

Casper, J., Zweig, A.S., Villarreal, C., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Karolchik, D., *et al.* (2018). The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res* *46*, D762-D769.

Cassart, C., Drogat, J., Migeot, V., and Hermand, D. (2012). Distinct requirement of RNA polymerase II CTD phosphorylations in budding and fission yeast. *Transcription* *3*, 231-234.

Chanfreau, G., Noble, S.M., and Guthrie, C. (1996). Essential yeast protein with unexpected similarity to subunits of mammalian cleavage and polyadenylation specificity factor (CPSF). *Science* *274*, 1511-1514.

Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. (2012). Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in *Saccharomyces cerevisiae*. *Eukaryot Cell* *11*, 417-429.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* *469*, 368-373.

Corden, J.L. (2013). RNA polymerase II C-terminal domain: Tethering transcription to transcript and template. *Chem Rev* *113*, 8423-8455.

Coulombe, C., Poitras, C., Nordell-Markovits, A., Brunelle, M., Lavoie, M.A., Robert, F., and Jacques, P.E. (2014). VAP: a versatile aggregate profiler for efficient genome-wide data representation and discovery. *Nucleic Acids Res* *42*, W485-493.

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

Descostes, N., Heidemann, M., Spinelli, L., Schuller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., *et al.* (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *Elife* *3*, e02105.

Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* *30*, 207-210.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* *113*, 8456-8490.

Exinger, F., and Lacroute, F. (1992). 6-Azauracil inhibition of GTP biosynthesis in *Saccharomyces cerevisiae*. *Curr Genet* *22*, 9-11.

Gelbart, M.E., Rechsteiner, T., Richmond, T.J., and Tsukiyama, T. (2001). Interactions of Isw2 chromatin remodeling complex with nucleosomal arrays: analyses using recombinant yeast histones and immobilized templates. *Mol Cell Biol* 21, 2098-2106.

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

Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev* 29, 849-861.

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

Guillemette, B., Bataille, A.R., Gevry, N., Adam, M., Blanchette, M., Robert, F., and Gaudreau, L. (2005). Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol* 3, e384.

Han, Z., Libri, D., and Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res* 45, 1355-1370.

Harlen, K.M., and Churchman, L.S. (2017). The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* 18, 263-273.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* 15, 2147-2158.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol Cell* 31, 925-932.

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

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* 7, 1534-1550.

Jasnovidova, O., Krejcikova, M., Kubicek, K., and Stefl, R. (2017). Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO Rep* 18, 906-913.

Jenny, A., Minvielle-Sebastia, L., Preker, P.J., and Keller, W. (1996). Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. *Science* 274, 1514-1517.

Jeronimo, C., Bataille, A.R., and Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chem Rev* 113, 8491-8522.

Jeronimo, C., Collin, P., and Robert, F. (2016a). The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain. *J Mol Biol* 428, 2607-2622.

Jeronimo, C., Langelier, M.F., Bataille, A.R., Pascal, J.M., Pugh, B.F., and Robert, F. (2016b). Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. *Mol Cell* 64, 455-466.

Jeronimo, C., and Robert, F. (2014). Kin28 regulates the transient association of Mediator with core promoters. *Nat Struct Mol Biol* 21, 449-455.

Jeronimo, C., Watanabe, S., Kaplan, C.D., Peterson, C.L., and Robert, F. (2015). The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations. *Mol Cell* 58, 1113-1123.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14, R36.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* 17, 1279-1286.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedeja, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517-522.

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

Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26, 1891-1896.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* 12, 283-294.

Laughery, M.F., Hunter, T., Brown, A., Hoopes, J., Ostbye, T., Shumaker, T., and Wyrick, J.J. (2015). New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. *Yeast* 32, 711-720.

Lenstra, T.L., Tudek, A., Clauder, S., Xu, Z., Pachis, S.T., van Leenen, D., Kemmeren, P., Steinmetz, L.M., Libri, D., and Holstege, F.C. (2013). The role of Ctk1 kinase in termination of small non-coding RNAs. *PLoS One* 8, e80495.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* 9, 1101-1111.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17, 1195-1201.

Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M., and Strathern, J.N. (2006). Mutations in the *Saccharomyces cerevisiae* RPB1 gene conferring hypersensitivity to 6-azauracil. *Genetics* 172, 2201-2209.

Malik, I., Qiu, C., Snavely, T., and Kaplan, C.D. (2017). Wide-ranging and unexpected consequences of altered Pol II catalytic activity in vivo. *Nucleic Acids Res* 45, 4431-4451.

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

Mayer, A., Heidemann, M., Lidschreiber, M., Schrieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336, 1723-1725.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* *17*, 1272-1278.

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

Mischo, H.E., and Proudfoot, N.J. (2013). Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta* *1829*, 174-185.

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

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol Cell* *29*, 577-587.

Nemec, C.M., Yang, F., Gilmore, J.M., Hintermair, C., Ho, Y.H., Tseng, S.C., Heidemann, M., Zhang, Y., Florens, L., Gasch, A.P., *et al.* (2017). Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner. *Proc Natl Acad Sci U S A* *114*, E3944-E3953.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* *497*, 127-131.

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

Porraa, O., and 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.

Qiu, H., Hu, C., and Hinnebusch, A.G. (2009). Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell* *33*, 752-762.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841-842.



Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., *et al.* (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306-2309.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev* 23, 1247-1269.

Roy, K., Gabunilas, J., Gillespie, A., Ngo, D., and Chanfreau, G.F. (2016). Common genomic elements promote transcriptional and DNA replication roadblocks. *Genome Res* 26, 1363-1375.

Schneider, B.L., Seufert, W., Steiner, B., Yang, Q.H., and Futcher, A.B. (1995). Use of polymerase chain reaction epitope tagging for protein tagging in *Saccharomyces cerevisiae*. *Yeast* 11, 1265-1274.

Schrieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L.A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol* 21, 175-179.

Schuller, R., Forne, I., Straub, T., Schrieck, A., Texier, Y., Shah, N., Decker, T.M., Cramer, P., Imhof, A., and Eick, D. (2016). Heptad-Specific Phosphorylation of RNA Polymerase II CTD. *Mol Cell* 61, 305-314.

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

Shah, N., Maqbool, M.A., Yahia, Y., El Aabidine, A.Z., Esnault, C., Forne, I., Decker, T.M., Martin, D., Schuller, R., Krebs, S., *et al.* (2018). Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals. *Mol Cell* 69, 48-61 e46.

Skaar, D.A., and Greenleaf, A.L. (2002). The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. *Mol Cell* 10, 1429-1439.

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

Soares, L.M., He, P.C., Chun, Y., Suh, H., Kim, T., and Buratowski, S. (2017). Determinants of Histone H3K4 Methylation Patterns. *Mol Cell* 68, 773-785 e776.

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.

Suh, H., Ficarro, S.B., Kang, U.B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Mol Cell* *61*, 297-304.

Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., and Lavin, M.F. (2009). Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum Mol Genet* *18*, 3384-3396.

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.

Tietjen, J.R., Zhang, D.W., Rodriguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S., *et al.* (2010). Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* *17*, 1154-1161.

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

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.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and 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*, 795-804.

Wagschal, A., Rousset, E., Basavarajaiah, P., Contreras, X., Harwig, A., Laurent-Chabalier, S., Nakamura, M., Chen, X., Zhang, K., Meziane, O., *et al.* (2012). Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* *150*, 1147-1157.

Werner, F., and Weinzierl, R.O. (2002). A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* *10*, 635-646.

Wong, K.H., Jin, Y., and Struhl, K. (2014). TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Mol Cell* 54, 601-612.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033-1037.

Zaborowska, J., Egloff, S., and Murphy, S. (2016). The pol II CTD: new twists in the tail. *Nat Struct Mol Biol* 23, 771-777.

## 2.3.7 Figures

Figure 2.1

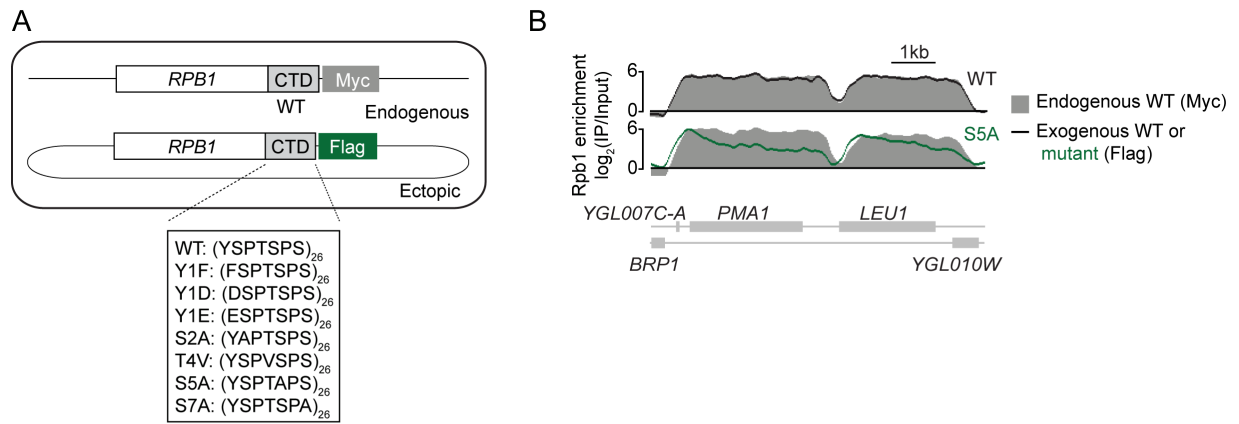


Figure 2.2

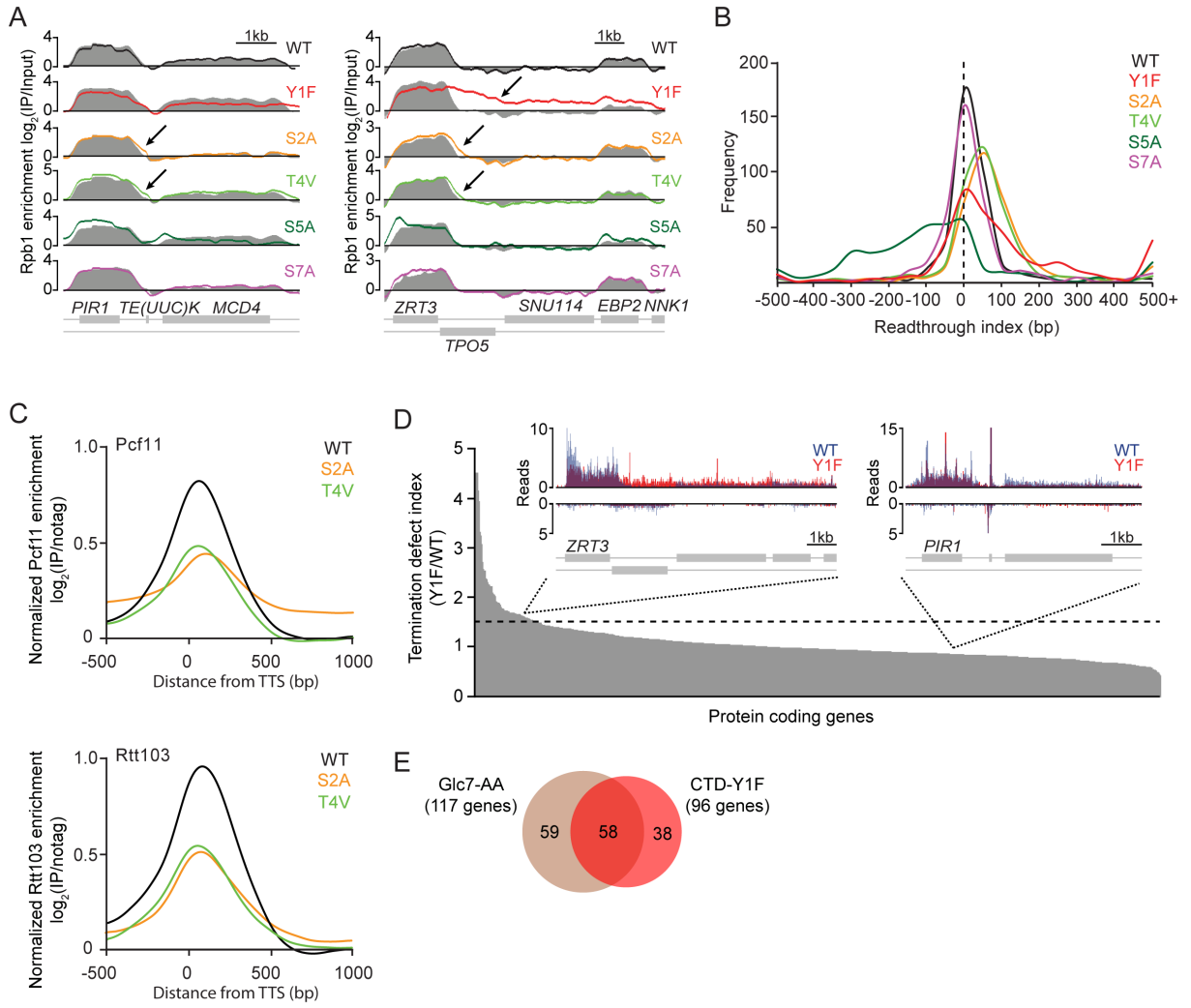


Figure 2.3

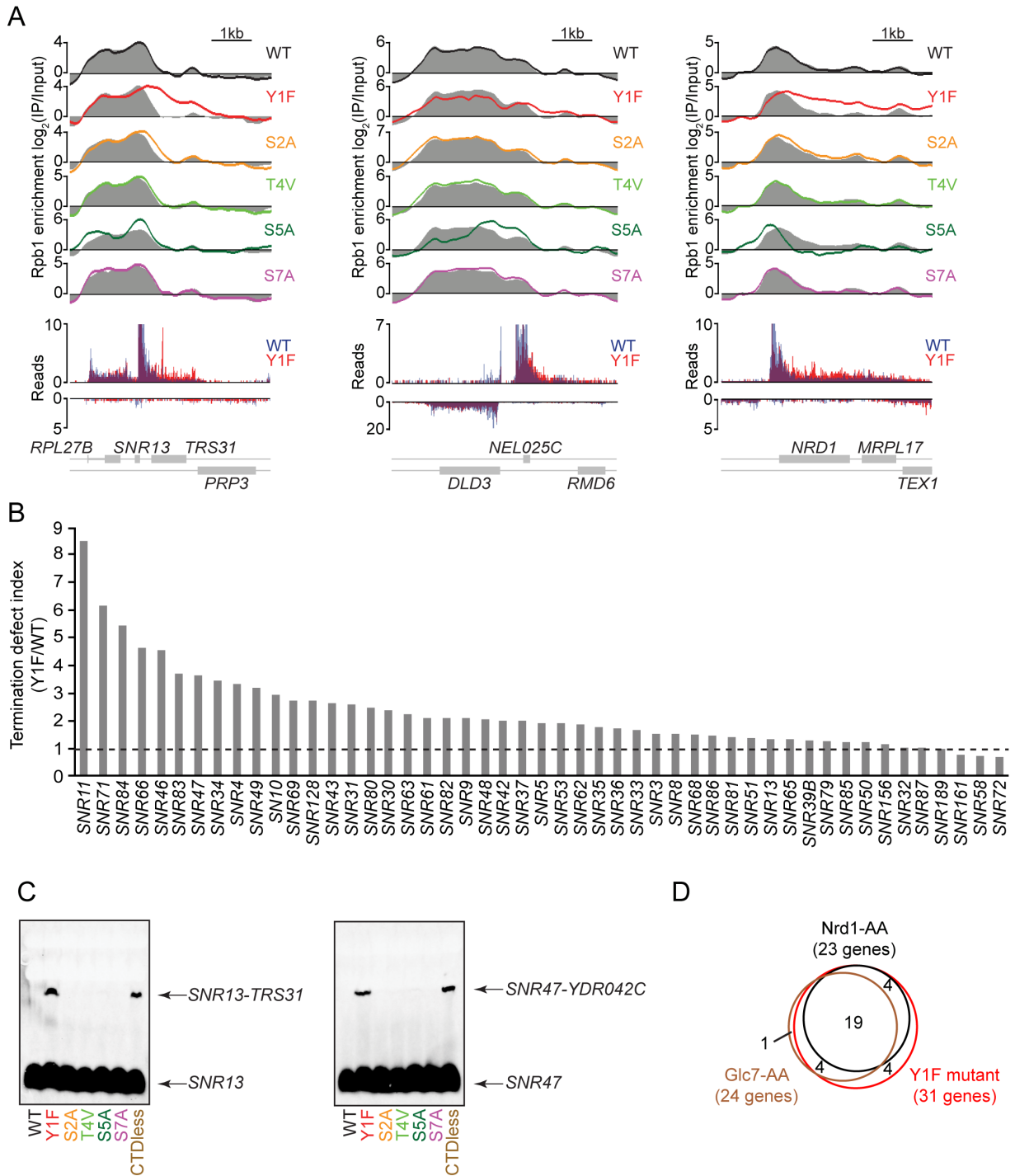


Figure 2.4

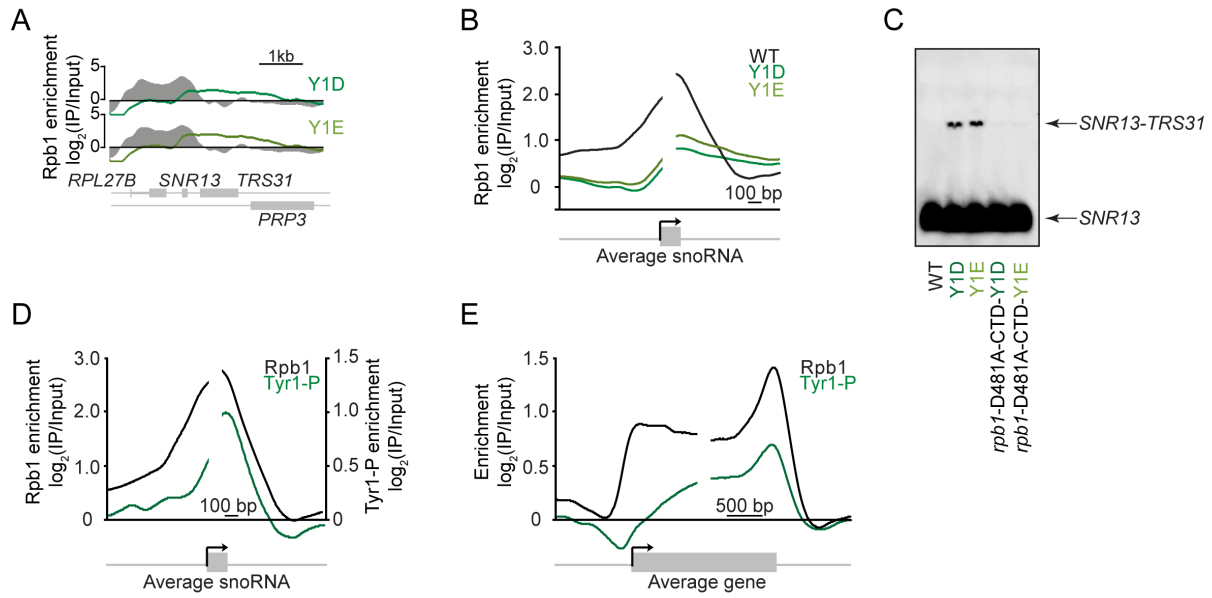


Figure 2.5

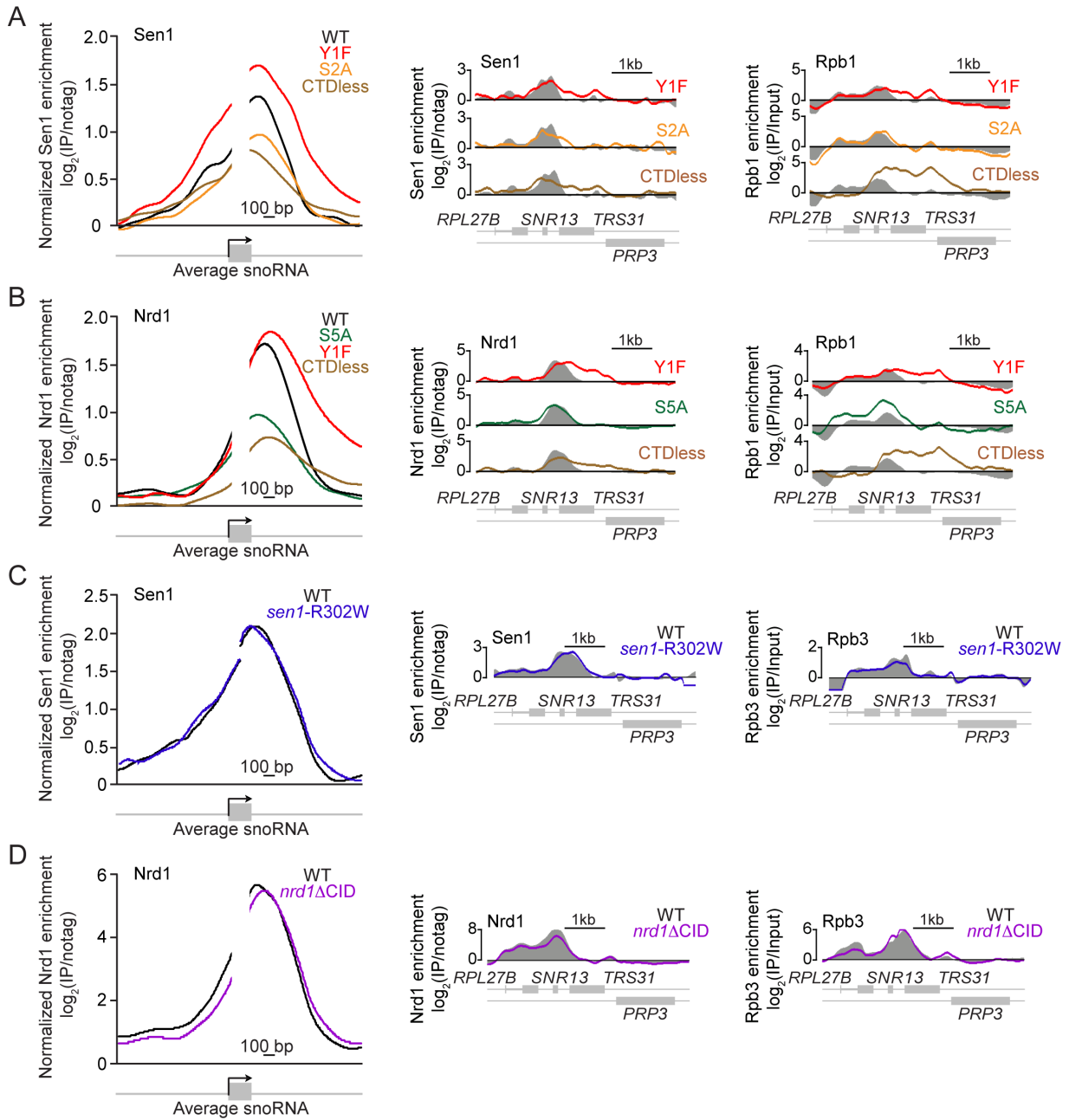




Figure 2.6

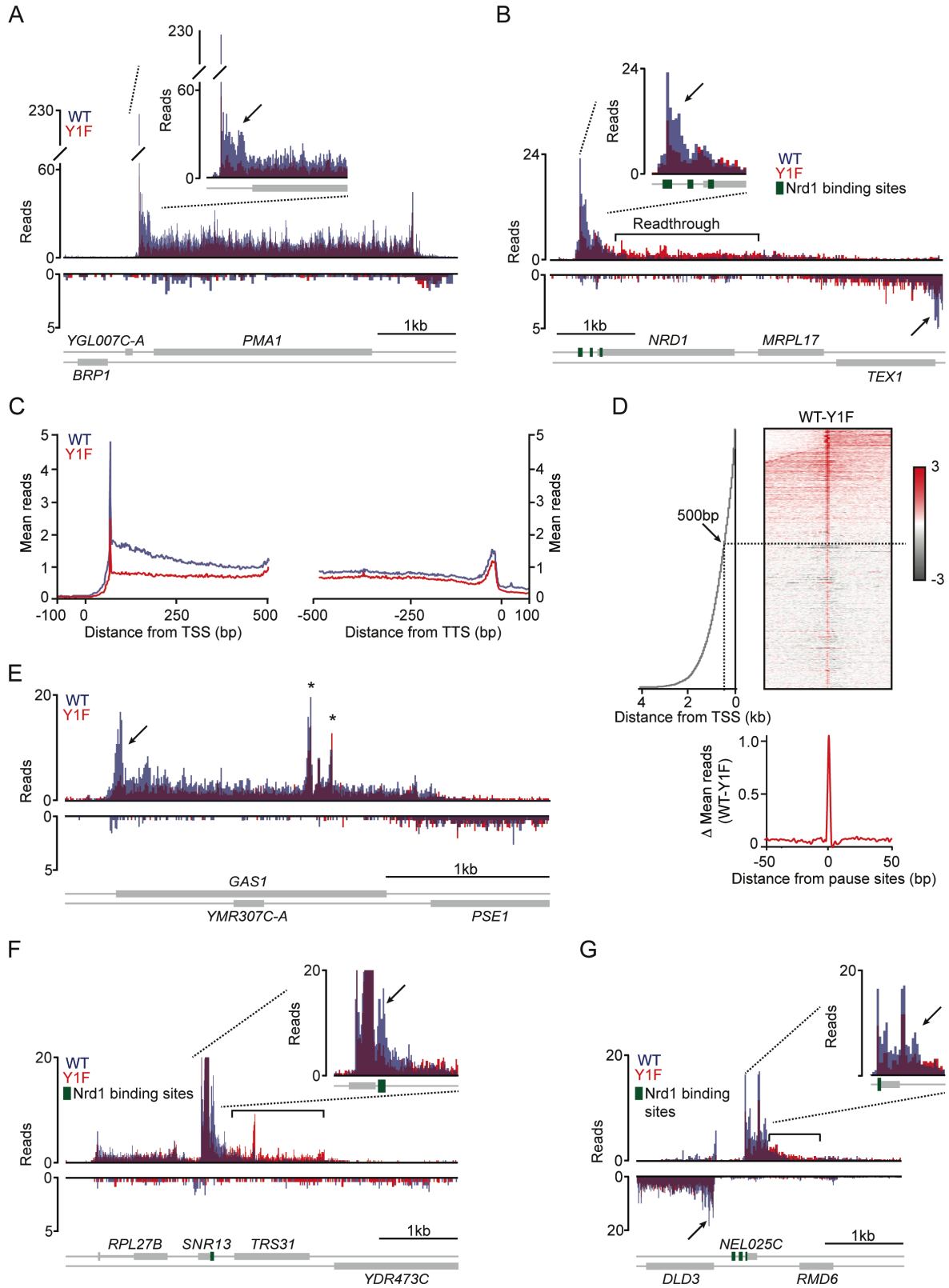
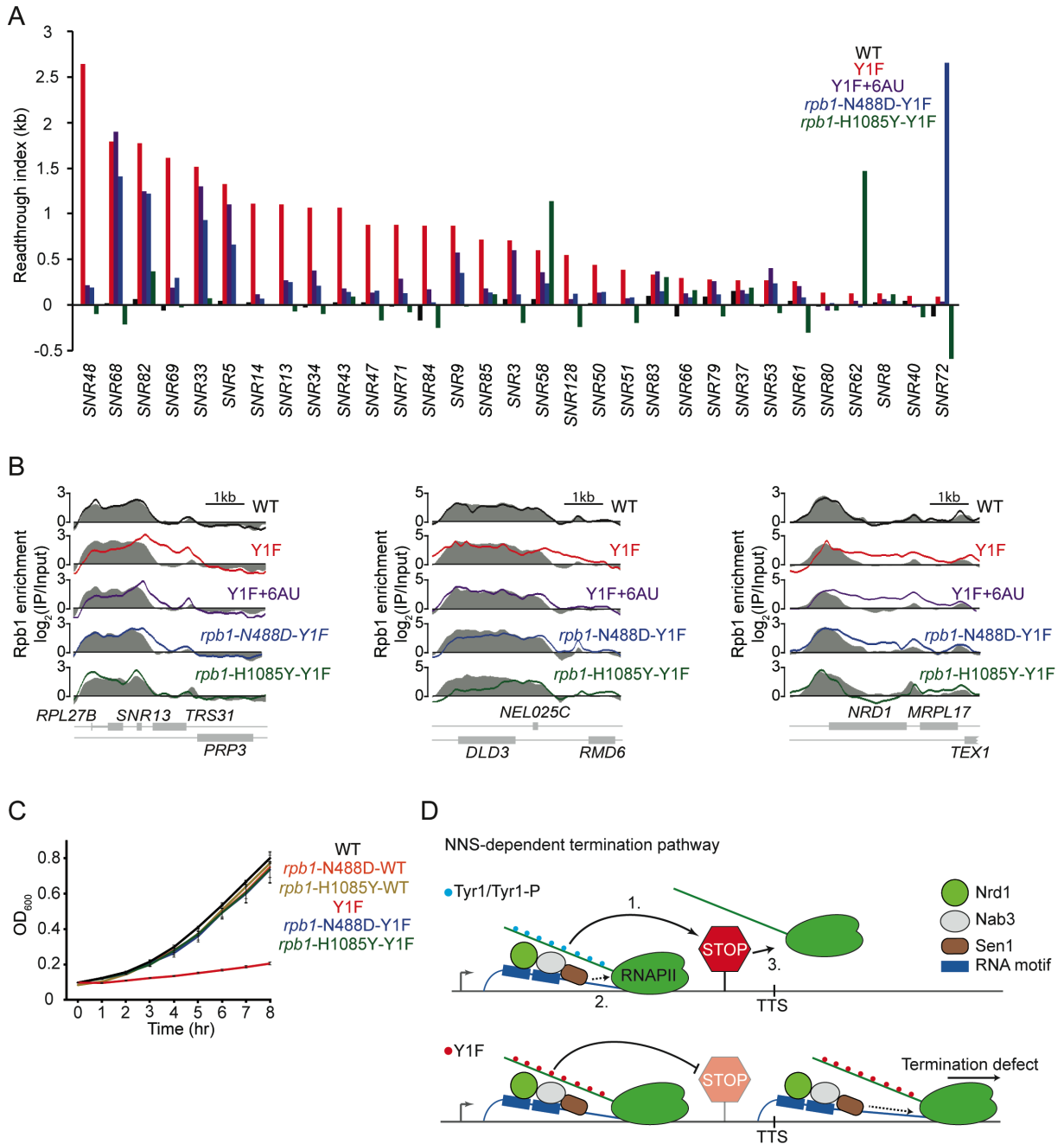


Figure 2.7



### 2.3.8 Figure legends

#### **Figure 2.1 A dual-tag system to explore the role of the RNAPII CTD in transcription.**

A) A Flag-tagged RPB1 allele, carrying different CTD mutations, is ectopically expressed in cells expressing the endogenous RPB1 WT, which is Myc-tagged. Mutations for phosphorylatable residues in all 26 CTD repeats were changed into non-phosphorylatable or phosphomimetic residues (see Table S6 for the exact sequence of each mutant). B) The enrichment of the endogenous WT Rpb1 (Myc-tagged) is shown as a grey shadow together with the enrichment of the exogenous (Flag-tagged) WT (black trace) or S5A mutant (green trace) Rpb1 over the PMA1 and LEU1 genes. See also Figure S1.

#### **Figure 2.2 CTD Ser2 and Thr4 are ubiquitously required for efficient termination at protein-coding genes, while Tyr1 has a gene-specific role.**

A) The enrichment of endogenous WT (grey shadow) and mutants (colored traces) RNAPII (Rpb1) is shown over the PIR1 (left) and ZRT3 (right) genes. Arrows indicate termination defects. B) Distribution of the readthrough index of expressed protein-coding genes (n=300, see STAR Methods) in various CTD mutants. Only genes with a downstream neighbour in the opposite direction were used to minimise effects from the downstream promoter. C) Pcf11 (top) and Rtt103 (bottom) average occupancy around the transcription termination site (TTS) of expressed protein-coding genes (n=625, see STAR Methods) in the S2A and T4V mutants after nuclear depletion of the endogenous WT Rpb1 protein. The occupancy of the termination factors is normalized to the level of Rpb1 at the TTS (See STAR Methods). See also Figures S2A and S2B for non-normalized profiles. D) Termination defect index, calculated from the NET-seq experiment, at protein-coding genes in Y1F cells for expressed protein-coding genes (n=712, see STAR Methods). The dashed line represents the threshold used (FDR=5%). NET-seq tracks for the ZRT3 and PIR1 genes are shown as insets. E) Venn diagram showing the overlap between expressed protein-coding genes (n=287) with a termination defect index more than 100 bp in the Y1F (n=96) and Glc7-AA (n=117) mutants. See also Figures S2 and S3.

**Figure 2.3 CTD Tyr1 is important for snoRNA termination.**

A) The enrichment of RNAPII endogenous WT (grey shadow) and mutants (colored traces) is shown over the SNR13 (left), NEL025C (middle) and NRD1 (right) genes. The bottom set of tracks shows WT (blue) and Y1F (red) NET-seq signal on both strands around these three genes. B) Termination defect index, calculated from the NET-seq experiment, at snoRNAs in Y1F cells (n=49, see STAR Methods). C) Northern blot showing the expression of the SNR13 (left) and SNR47 (right) transcripts from WT and various CTD RNAPII mutants after depletion of the endogenous WT Rpb1 protein by anchor-away. Experiments were done in *rrp6Δ* cells in order to stabilize transcripts. D) Venn diagram showing the overlap between snoRNAs with a readthrough index longer than 100 bp in Nrd1-AA, Glc7-AA and Y1F cells based on ChIP-chip experiments. See also Figure S4.

**Figure 2.4 Tyr1 phosphomimetics disrupt snoRNA termination.**

A) The enrichment of endogenous WT RNAPII (grey shadow) and RNAPII carrying phosphomimetic Tyr1 CTD mutations (colored traces) is shown over the SNR13 gene. B) Rpb1 WT and phosphomimetic Tyr1 CTD mutants average occupancy over snoRNA loci (n=33, see STAR Methods). C) SNR13 transcripts from WT RNAPII and RNAPII carrying phosphomimetic Tyr1 CTD mutations were analyzed by Northern blot in *rrp6Δ* cells after nuclear depletion of the endogenous Rpb1. The blot also includes transcripts from phosphomimetic CTD mutants that also carry a mutation in the active site (D481A) abrogating the activity of the polymerase. D) Rpb1 and CTD Tyr1-P average occupancy over snoRNA loci (n=33, see STAR Methods). E) Rpb1 and CTD Tyr1-P average occupancy over expressed protein-coding genes (n=625, see STAR Methods).

**Figure 2.5 Nrd1 and Sen1 are recruited to snoRNA loci independently of their interaction with the CTD.**

A, B) Left, Sen1 (A) and Nrd1 (B) average occupancy over snoRNA loci (n=33, see STAR Methods) in the indicated CTD mutants after nuclear depletion of the endogenous Rpb1. Termination factor occupancy was normalized to Rpb1 occupancy (see STAR Methods and

Figure S5 for non-normalized profiles). Right, the enrichment of Sen1 (A), Nrd1 (B) and RNAPII is shown in WT (grey shadows) and the indicated CTD mutants (colored traces) over the SNR13 locus after nuclear depletion of the endogenous Rpb1. C, D) Left, Sen1 (C) and Nrd1 (D) average occupancy over snoRNA loci (n=33, see STAR Methods) in WT (black) and sen1-R302W (C, blue) or nrd1 $\Delta$ CID (D, purple) cells. Sen1 and Nrd1 occupancies were normalized to Rpb3 occupancy (see STAR Methods and Figure S5 for non-normalized profiles). Right, the enrichment of Sen1 (C), Nrd1 (D) and RNAPII is shown in WT (grey shadow) and sen1-R302W (C, blue) or nrd1 $\Delta$ CID (D, purple) cells over the SNR13 locus. See also Figure S5.

**Figure 2.6 Pausing in the 5' end of genes is decreased in Y1F cells.**

A-B) Genome browser views of WT (blue) and Y1F (red) NET-seq signals over the (A) PMA1 and (B) NRD1 genes. The overlap between the WT and the Y1F signal is shown in purple. Brackets indicate readthrough of RNAPII. Inlets show zoom in views of the 5' end of genes, emphasising the pausing events (arrows) present in WT and absent in the Y1F mutant. The green boxes represent Nrd1 binding sites based on (Creamer et al., 2011). C) WT (blue) and Y1F (red) average NET-seq signal over expressed protein-coding genes (n=712, see STAR Methods). D) Heat map representation of the difference in normalized NET-seq signal between Y1F and WT cells around pause sites ordered by their distance from the TSS (shown on the left). The average signal is also shown as a trace at the bottom. E) Genome browser view of WT (blue) and Y1F (red) NET-seq signals over the GAS1 gene. Arrow indicates a pausing event at the 5' end and asterisks indicate internal pause sites. F-G) Genome browser views of WT (blue) and Y1F (red) NET-seq signals over the snoRNA SNR13 (F) and the CUT NEL025C (G). The green boxes represent Nrd1 binding sites based on (Creamer et al., 2011). See also Figure S6.

**Figure 2.7 Impairing elongation rescues the termination defects of Y1F cells.**

A) Readthrough index at Tyr1-dependent snoRNAs in WT, Y1F, Y1F treated with 6AU, rpb1-N488D-Y1F and rpb1-H1085Y-Y1F cells. B) The enrichment of endogenous WT (grey shadow) and mutants (colored traces) RNAPII is shown over the SNR13 (left), NEL025C

(middle) and NRD1 (right) locus. C) Growth curves of WT and slow *rpb1* mutants, alone or in combination with a CTD Y1F mutation, during 8 hr in YNB –HIS media. D) Proposed model for the role of Tyr1 in termination by the NNS pathway. See also Figure S7.

### 2.3.9 Star methods

Detailed methods are provided in the online version of this paper and include the following:  
Outline to be created by the Production team.

#### 2.3.9.1 Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Mouse monoclonal anti-Flag (M2)	Sigma-Aldrich	Cat#F3165; RRID: AB_259529
Mouse monoclonal anti-Myc (9E10)	This study	N/A
Rabbit polyclonal anti-Rpb1 ( $\gamma$ -80; N-term)	Santa Cruz Biotechnology	Cat#sc-25758; RRID: AB_655813
Mouse monoclonal anti-Rpb3 (1Y26)	Neoclone; BioLegend	Cat#W0012, RRID: AB_2564529; Cat#665003; RRID: AB_2565004; Cat#665004; RRID: AB_2565221
Rat anti-phospho-RNAPII (Tyr1) (3D12)	D. Eick	NA
IRDye 680RD Donkey anti-mouse	Li-Cor Biosciences	Cat#lic-926-68072
IRDye 800CW Donkey anti-rabbit	Li-Cor Biosciences	Cat#lic-926-32213
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
5-Fluoroorotic acid	Toronto Research Chemicals	F595000
Rapamycin	Bio Basic	R706203
6-azauracil	Sigma-Aldrich	A1757-5G
Dynabeads Protein G	Thermo Fisher Scientific	10004D
Dynabeads Pan Mouse IgG	Thermo Fisher Scientific	11042
Proteinase K (20 mg/mL)	Thermo Fisher Scientific	25530049
RNase A (50 mg)	Sigma-Aldrich	R6513-50mg
Glycogen (20 mg/mL)	Roche	10901393001
5-(3-aminoallyl)-dUTP (aa-dUTP)	Thermo Fisher Scientific	AM8439
NEB2 10X buffer	New England Biolabs	B7002S
T4 DNA polymerase (3 U/uL)	New England Biolabs	M0203L
T4 DNA ligase (5 U/uL)	Life	15224041
5-(3-aminoallyl)-dUTP	Thermo Fisher Scientific	AM8439
10X Thermopol buffer	New England Biolabs	B9004S
Taq DNA polymerase	Life	18038240
Pfu DNA polymerase (2.5 U/uL)	Fisher	EP0502
Cy3 mono-reactive dye pack	GE Healthcare	CA95017-373L
Cy5 mono-reactive dye pack	GE Healthcare	CA95017-379L

Klenow Fragment (3'→5' exo-)	New England Biolabs	M0212M
DyLight800 NHS-Ester	Thermo Fisher Scientific	46422
Random primers (3 ug/uL)	Life	48190011
40% glyoxal solution	Sigma-Aldrich	50649-25ML
501-X8 (D) ion exchange	BioRad	143-6425
RQ1 RNase-free DNase I	Promega	MG101
3xFlag peptide	Sigma-Aldrich	F4799-4MG
T4 RNA Ligase 2, truncated K227Q	New England Biolabs	M0351S
SYBR gold	Invitrogen	S11494
Costar Spin-X centrifuge tube filter	VWR	29442-756
Superaise.In	Ambion	AM2694
Superscript IV	Invitrogen	18090050
CircLigase ssDNA Ligase	Mandel	LGN-CL4111K
Phusion DNA polymerase	New England Biolabs	M0530S
<b>Critical Commercial Assays</b>		
QIAquick PCR purification kit	Qiagen	28106
Qiagen miRNeasy mini kit	Qiagen	217004
<b>Deposited Data</b>		
Raw and analyzed data	This study	GEO: GSE115028
RNAPII occupancy after Glc7 anchor-away	Schrieck et al., 2014	ArrayExpress: E-MTAB-1528
<i>rpb1</i> -E1103G occupancy	Soares et al., 2017	GEO: GSE95356
<b>Experimental Models: Organisms/Strains</b>		
<i>Saccharomyces cerevisiae</i> : Strain background: W303	Laboratory of Rodney Rothstein	ATCC-208352
<i>Saccharomyces cerevisiae</i> : Strain background: BY4741	Laboratory of Jef D. Boeke	ATCC-201388
Please see Table S1 for derivative strains		
<b>Oligonucleotides</b>		
Please see Table S4 for all oligonucleotides		
<b>Recombinant DNA</b>		
Please see Table S2 for all plasmids		
<b>Software and Algorithms</b>		
Versatile Aggregate Profiler (version 1.1.0)	Brunelle et al., 2015; Coulombe et al., 2014	<a href="http://lab-jacques.recherche.usherbrooke.ca/vap/">http://lab-jacques.recherche.usherbrooke.ca/vap/</a>
UCSC Genome Browser	Casper et al., 2018	<a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a>
TopHat2	Kim et al., 2013	<a href="http://ccb.jhu.edu/software/tophat/index.shtml">http://ccb.jhu.edu/software/tophat/index.shtml</a>
Bamtools	(Barnett et al., 2011)	<a href="https://github.com/pmezmer31/bamtools">https://github.com/pmezmer31/bamtools</a>
ucsc_tools	N/A	<a href="http://hgdownload.soe.ucsc.edu/downloads.html#source_downloads">http://hgdownload.soe.ucsc.edu/downloads.html#source_downloads</a>



In-house tool to trim NET-seq reads, keeping only the first sequenced base	This study	<a href="https://github.com/IRCM/bed-tools-j/releases/tag/1">https://github.com/IRCM/bed-tools-j/releases/tag/1</a>
BEDTools (version 2.19.1)	Quinlan and Hall, 2010	<a href="http://bedtools.readthedocs.io/en/latest/">http://bedtools.readthedocs.io/en/latest/</a>
Samtools (version 1.2)	(Li et al., 2009)	<a href="http://www.htslib.org/">http://www.htslib.org/</a>
Picard Tools (version 1.56.0)	N/A	<a href="http://broadinstitute.github.io/picard/">http://broadinstitute.github.io/picard/</a>

### Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, François Robert ([francois.robert@ircm.qc.ca](mailto:francois.robert@ircm.qc.ca)).

### 2.3.9.2 Experimental model and subject details

Yeast (*S. cerevisiae*) strains used in this study are described in **Table S4**. Yeast cells carrying no plasmid were grown at 30°C in YPD (Yeast extract peptone with 2% glucose) medium, supplemented with 44 µM adenine. Cells expressing *RPB1* alleles on plasmids were grown at 30°C in Yeast Nitrogen Base (YNB) medium lacking histidine and supplemented with 2% glucose. Typically, one colony was used to inoculate 5 mL of appropriate media for pre-culture grown overnight. Cell density was measured with a spectrophotometer at OD<sub>600</sub> and cells were inoculated at OD<sub>600</sub>=0.1 in 50 mL, 100 mL or 1L of appropriate media (depending on the experiment, see below) and grown at 30°C with agitation until OD<sub>600</sub> reached 0.7-0.8.

For the anchor-away experiments, cells were treated with 1 µg/mL of rapamycin for 90 min before cross-linking. For the 6-azauracyl (6AU) experiments, cells were grown in YNB medium lacking histidine and uracil (-HIS, -URA) and treated with 50 µg/mL of 6AU for 2 hr before cross-linking. All treatments (6-AU or rapamycin) were performed at OD<sub>600</sub>=0.5.

### 2.3.9.3 Detailed methods

#### Yeast strains and plasmids

Yeast (*S. cerevisiae*) strains used in this study are described in **Table S4**. Epitope tagging was performed by standard DNA recombination methods using the following plasmids: p3FLAG-KanMX (Gelbart et al., 2001), pMPY-3xMYC (Schneider et al., 1995) and pFA6a-FRB-KanMX6 (Haruki et al., 2008). The *sen1*-R302W mutation was introduced by CRISPR/Cas9 genome editing using a protocol slightly modified form ((Laughery et al., 2015); see below). To construct the *rpb1*-CTD-mutants-3xFlag plasmids (listed in **Table S5**), various CTD mutants (where all residues of interest were replaced by an homologous amino acid that do not have the hydroxyl group required for phosphorylation (Y1F, S2A, T4V, S5A, S7A) or by residues that mimic a phosphorylation event (Y1D and Y1E); see **Table S6** for complete sequences) were synthesized as minigenes by Integrated DNA Technologies or by GenScript and cloned into pFR467, a derivative of pRS313 (-HIS) containing a CTD-less *rpb1*-3xFlag under control of *RPB1* promoter and terminator. Point mutations in *RPB1* conferring reduced catalytic rates (H1085Y and N488D) were introduced into pFR482 and pFR511 by inverted PCR to generate pFR683, pFR684, pFR686 and pFR687 (see **Table S5**). Finally, the catalytic dead *rpb1* mutation (D481A) (Werner and Weinzierl, 2002) was introduced into pFR574 and pFR576 by inverted PCR to generate pFR728 and pFR729 (see **Table S5**). Each construction was validated by sequencing and by immunoblotting.

#### Assays of RNAPII expression and stability

Cells expressing the endogenous *RPB1*-WT-3xMyc allele and any of the exogenous *rpb1*-CTD-mutant-3xFlag alleles were inoculated at OD<sub>600</sub>=0.1 in 10 mL YNB -HIS supplemented with 2% glucose and grown at 30°C. When cells reached OD<sub>600</sub>=0.8, cells were washed with ice-cold Tris-buffered saline (TBS; 50 mM Tris-HCl pH 7.5, 150 mM NaCl), resuspended in 700 µL of Lysis buffer (50 mM HEPES-KOH 1 M pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Na-deoxycholate) supplemented with protease inhibitors (1 mM PMSF, 1 mM benzamidine, 10 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin) and lysed by bead beating. Lysates were centrifuged 5 min at 4°C and 10 µL of 5X

Laemmli buffer were added to 40  $\mu$ L of supernatant. Samples were boiled 5 min, separated on SDS-PAGE gels and transferred on nitrocellulose membranes. Membranes were incubated in 5% (w/v) milk or BSA in TBS supplemented with 0.1% Tween-20 (TBST) containing primary antibody (anti-Flag M2 (1:2,000), anti-Myc 9E10 (1:2,000), anti-Rpb1 y-80 (1:100) or anti-Spt16 (1:1,000)) overnight at 4°C. Membranes were washed with TBST and incubated for 30 minutes with secondary antibody coupled to IRDye (LI-COR Biosciences), diluted in TBST plus 5% (w/v) milk. Membranes were washed with TBST and scanned on the Odyssey infrared imaging system (LI-COR Biosciences).

### **Serial-dilution growth assay**

Cells expressing the endogenous *RPBI*-WT-3xMyc allele and any of the exogenous *rpb1*-CTD-mutant-3xFlag alleles were grown to saturation, washed with H<sub>2</sub>O and diluted in water to OD<sub>600</sub>=8 followed by 10 fold serial dilutions. Cell dilutions were then spot on YNB – HIS plates supplemented with 2% glucose, incubated at 30°C and inspected daily.

### **Growth curves**

Cells expressing the endogenous *RPBI*-WT-3xMyc allele and the indicated exogenous *rpb1*-CTD-mutant-3xFlag alleles were incubated in 5 mL of YNB –HIS supplemented with 2% glucose and grown overnight. Cells were then inoculated at OD<sub>600</sub>=0.1 in 50 mL of YNB –HIS supplemented with 2% glucose, incubated at 30°C and 1 mL of cultured cells was used to determine OD<sub>600</sub> each hr for a total of 8 hr.

### **RNA extraction**

Cells (*rrp6* $\Delta$ ) expressing an endogenous *RPBI*-WT-FRB allele and any of the exogenous *rpb1*-CTD-mutant-3xFlag alleles were inoculated at OD<sub>600</sub>=0.1 and grown at 30°C in YNB –HIS supplemented with 2% glucose. When OD<sub>600</sub> reached 0.5, cells were treated with 90  $\mu$ g/mL rapamycin for 90 min. Cells were harvested and washed in 5 mL of H<sub>2</sub>O. Dried cell pellets were snapped-frozen in liquid nitrogen. Total RNA was extracted using the hot phenol extraction method. In brief, cell pellets were resuspended in 400  $\mu$ L of hot acid

phenol:chloroform (5:1) and 400  $\mu$ L of TES solution (10 mM Tris-HCl pH 7.5, 10 mM EDTA and 0.5% SDS) was immediately added to the samples, which were vortexed vigorously. Samples were incubated 60 min at 65°C and vortexed every 10 min. Samples were centrifuged 20 min at room temperature at maximum speed. The aqueous phase was transferred to a clean microcentrifuge tube for a second acid phenol:chloroform (5:1) extraction followed by an extraction with chloroform. DNA was ethanol-precipitated, washed with ice-cold 70% ethanol, resuspended in 200  $\mu$ L of DEPC-H<sub>2</sub>O and stored at -80°C.

### **Northern blots**

Northern blots were performed as previously described (Jeronimo et al., 2015) with minor modifications. To prepare probes, regions of interest were PCR-amplified from yeast genomic DNA using primers from (Kim et al., 2009) for *SNR13* and from (Tudek et al., 2014) for *SNR47*. PCR products were purified using QIAquick PCR purification kit (Qiagen, 28106) following the manufacturer's protocol. 100 ng of DNA obtained by PCR was mixed with 20  $\mu$ L of 2.5X Random primer mix (125 mM Tris pH 6.8, 12.5 mM MgCl<sub>2</sub>, 0.5  $\mu$ g/ $\mu$ L Random primers (Life, 48109911), 25 mM  $\beta$ -mercaptoethanol and 0.5 mg/mL BSA) and incubated 5 min at 95°C. Samples were placed on ice for 10 min and mixed with 5  $\mu$ L of 10X NEB2 buffer, 5  $\mu$ L of aa-dUTP-dNTP mix (5mM each dATP, dGTP and 5mM dCTP, 3mM dTTP, 2mM aminoallyl-dUTP) and 25 units of Klenow (NEB, M0212M), and incubated overnight at 37°C. DNA was purified using QIAquick PCR purification kit following the manufacturer's protocol with the following modifications: Wash buffer was replaced by Phosphate wash buffer (5 mM KPO<sub>4</sub> pH 8.5 and 80% ethanol) and EB buffer by Phosphate elution buffer (4 mM KPO<sub>4</sub> pH 8.5). Eluates were dried in a speed-vac and resuspended in 4.5  $\mu$ L of freshly prepared 0.1 M Na<sub>2</sub>CO<sub>3</sub> pH 9.0. 4.5  $\mu$ L of DyLight800 NHS-Ester dye (Thermo Fisher Scientific, 46422) was added and probes were incubated 1 hr in the dark at room temperature. Probes were purified using QIAquick PCR purification kit following the manufacturer's protocol. Probes were quantified on a Nanodrop and 5  $\mu$ L was run on 1% agarose gel, stained with ethidium bromide and scanned with an Odyssey infrared imaging system (LI-COR Biosciences) in the 800 nm channel to insure proper dye incorporation.

50 µg of total RNA was dried using a speed-vac and resuspended in 2 µL of DEPC-H<sub>2</sub>O. 10 µL of glyoxal mix (prepared by mixing 6 mL DMSO, 2 mL deionized glyoxal, 1.2 mL 10X BPTE and 0.6 mL 80% glycerol) was added to the samples, which were mixed and incubated 1 hr at 55°C for denaturation. Samples were placed on ice for 10 min, centrifuged and load on a 1% agarose BPTE (10 mM PIPES, 30 mM Bis-Tris, 1 mM EDTA pH 8.0) gel and run at 5 V/cm for 150 min. The gel was transferred on Hybond N+ nylon membrane overnight by capillarity. The next day, the membrane was cross-linked using a Stratalinker 2400 UV Crosslinker (BioSurplus), stained with ethylene blue and wash with H<sub>2</sub>O. Finally, the membrane was destained in pre-warmed 1X SSC, 0.1% SDS at 65°C for at least 20 min. The membrane was incubated for 4 hr at 65°C in 15 mL of pre-warmed pre-hybridization solution (6X SSC, 5X Denhardts (0.1% BSA, 0.1% Ficoll 400, 0.1% polyvinylpyrrolidone)). 100 ng of probes were denatured 10 min at 95°C and placed on ice for 10 min before being added to the membrane, which was incubated overnight at 65°C within a sealed bag with agitation. The membrane was washed 3 times in 1X SSC, 0.1% SDS at 65°C for 20 min and then washed three times at room temperature with the same buffer. Membranes were scanned with the Odyssey infrared imaging system (LI-COR Biosciences) in the 800 nm channel.

### **Point mutation using CRISPR**

The *senI*-R302W mutation was introduced directly in the genome by CRISPR/Cas9 genome editing using a protocol slightly modified form (Laughery et al., 2015). Briefly, a pair of oligonucleotides corresponding to a gRNA targeting a region around R302 (see **Table S7**) was designed for cloning into pML104 using <http://wyrickbioinfo2.smb.wsu.edu/crispr.html> generating pFR622 (see **Table S5**). Oligonucleotides were annealed and cloned into pML104 using *Swa*I and *Bcl*II restriction enzymes. Yeast cells were transformed with 400 ng of the gRNA-expressing plasmid (pFR622) and 300 pmol of annealed repair oligos (see **Table S7**) containing the desired mutations (including one silent mutation destroying the protospacer adjacent motif and one creating the R302W mutation) and plated on YNB –URA plates supplemented with 2% glucose. Genomic DNA from yeast clones was extracted and DNA over the target region was amplified by PCR and screened by sequencing. The gRNA-expressing plasmid was then ejected from the selected clones by plating on YNB

supplemented with 2% glucose and 1 mg/ml 5-Fluoroorotic acid. Clones were re-sequenced to confirm the presence of the desired point mutation.

### **ChIP-chip experiments**

ChIP-chip experiments were performed as previously described (Jeronimo et al., 2016b). Cells were inoculated from an overnight pre-culture at  $OD_{600}=0.1$  in the appropriate media (see above) and grown at 30°C until  $OD_{600}$  reached 0.7-0.8. 50 mL of culture was used for each experiment, excepted for the recruitment experiments, where 100 mL was used. Cells were cross-linked with 1% of formaldehyde for 30 min at room temperature and quenched with 125 mM glycine. Cells were washed twice with ice-cold TBS and snapped-frozen. Cell pellets were resuspended in 700  $\mu$ L of Lysis buffer (50 mM HEPES-KOH 1 M pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Na-deoxycholate) supplemented with protease inhibitors (1 mM PMSF, 1 mM benzamidine, 10  $\mu$ g/mL aprotinin, 1  $\mu$ g/mL leupeptin and 1  $\mu$ g/mL pepstatin) and lysed by bead beating. The extracts were sonicated with a Model 100 Sonic dismembrator equipped with a micro probe (Fisher Scientific), 4 x 20 sec at output 7 Watts, with a 1 min break between sonication cycles to obtain fragments averaging 500 bp. Samples were centrifuged at 20,000 g for 5 min at 4°C to remove cell debris and 500  $\mu$ L of soluble chromatin were used per ChIP experiment. 5  $\mu$ L (1% of the amount used for IP) was also kept as an Input sample. 50  $\mu$ L of Dynabeads coated with Pan Mouse IgG antibodies (Thermo Fisher Scientific, 11042) or Protein G (Thermo Fisher Scientific, 10004D), pre-coupled with the appropriate antibody (anti-Myc (10  $\mu$ L per IP), anti-Flag (Sigma-Aldrich F3165, 5  $\mu$ g per IP), anti-Rpb3 (Neoclone W0012, 3  $\mu$ L; BioLegend 665004, 1.5  $\mu$ L) or anti-phospho-RNAPII (Tyr1 3D12,5  $\mu$ L per IP)), was added to the chromatin extract which was incubated overnight at 4°C with agitation. Beads were washed twice with Lysis buffer, twice with Lysis buffer + 360 mM NaCl, twice with Wash buffer (10 mM Tris-HCl pH 8, 250 mM LiCl, 0.5 % NP40, 0.5 % Na-deoxycholate and 1 mM EDTA) and once with TE using a magnetic stand. Beads and Input samples were resuspended in 50  $\mu$ L of TE + 1% SDS and incubated overnight at 65°C. Supernatants were recovered and treated with RNase A (345  $\mu$ L TE (10 mM Tris-HCl pH 8, 1 mM EDTA), 3  $\mu$ L RNase A 10 mg/mL, 2  $\mu$ L Glycogen 20 mg/mL) at 37°C for 2 hr and subsequently subjected to Proteinase K (15  $\mu$ L 10% SDS, 7.5  $\mu$ L

Proteinase K 20 mg/mL) digestion at 37°C for 2 hr. DNA was extracted twice with phenol/chloroform/isoamyl alcohol, ethanol-precipitated and resuspended in 50 µL TE.

DNA blunting was done starting from 40 µL of ChIP DNA or about 400 ng of Input DNA. Samples were added to 11 µL 10X NEB2 buffer, 0.5 µL 10 mg/mL BSA, 0.5 µL 20 mM dNTPs and 0.2 µL (0.6 units) of T4 DNA pol (New England Biolabs, M0203L) in a final volume of 70 µL and incubated 20 min at 12°C. Then, 11.5 µL of 3 M NaOAc pH 5.2 and 2 µL of 20 mg/mL glycogen were added, followed by a phenol/chloroform/isoamyl alcohol extraction ethanol precipitation. DNA pellets were resuspended in 25 µL H<sub>2</sub>O and DNA linkers were ligated using 10 µL 5x ligase buffer, 6.7 µL annealed linkers (made of the annealing of the following oligos: oligo 1, GCGGTGACCCGGGAGATCTGAATTC and oligo 2, GAATTCAGATC) and 0.5 µL (2.5 units) of T4 DNA ligase (ThermoFisher, 15224041), overnight at 16°C. Samples were precipitated with 6 µL of 3M NaOAc pH 5.2 and ethanol, and resuspended in 25 µL H<sub>2</sub>O. 15 µL of labeling mix (4 µL 10X ThermoPol buffer, 2 µL of 5 mM a.a.-dUTP/dNTP mix (5mM each dATP, dGTP and dCTP, 3mM dTTP and 2mM aminoallyl-dUTP) and 1.25 µL of oligo 1 (from 40 µM stock) was add to each sample. Samples were placed in a thermocycler and temperature was raised to 55°C. 10 µL of enzyme mix (8 µL H<sub>2</sub>O, 1 µL 10X ThermoPoll buffer, 1 µL (5 units) Taq polymerase (ThermoFisher, 18038240) and 0.01 µL (0.025 units) Pfu (Fermentas, EP0502)) was added and the following program was run. 1) 72°C, 5 min, 2) 95°C, 2 min, 3) 31 cycles: 95°C, 30sec; 55°C, 30sec; 72°C, 1min, 4) 72°C, 4min, 5) 4°C, forever. PCR products were purified using QIAGEN PCR purification kit with few modifications: PE buffer was replaced by a Phosphate wash buffer (5 mM KPO<sub>4</sub> pH 8.5 and 80% ethanol) and EB buffer was replace by Phosphate elution buffer (4 mM KPO<sub>4</sub> pH 8.5). Supernatants were dried in a speed-vac and pellets were resuspended in 4.5 µL of freshly prepared 0.1 M Na<sub>2</sub>CO<sub>3</sub> pH 9.0 buffer. 4.5 µL of Cy5 (ChIP samples) or Cy3 (control samples)-ester dye (GE Healthcare, CA95017-373L and CA95017-379L) were added. Samples were incubated 1 hr at room temperature. 35 µL of 0.1 M NaOAc pH 5.2 was added and samples were purified on QIAGEN PCR purification kit (28106) using the manufacturer's protocol. ChIP and control samples were combined and dried in a speed-vac. Combined samples were resuspended in 110 µL of a hybridization mix (5 µL of 10 mg/mL ssDNA, 5 µL of 8 mg/mL tRNA and 100 µL DIG EASY (Roche, 11603558001)) and incubated 5 min at

95°C followed by 5 min at 42°C. Samples were incubated overnight on Agilent microarrays in SureHyb chambers with rotation. Microarrays were washed 5 min at room temperature with Wash I buffer (6X SSPE, 0.005% N-lauroylsarcosine) and washed a second time with a pre-warmed (42°C) Wash II buffer (0.06X SSPE) during 5 min at room temperature. Microarrays were scanned using an InnoScan900 (Innopsys) at 2µm resolution. The microarrays, described before (Jeronimo and Robert, 2014), were purchased from Agilent Technologies and contain a total of 180,000 Tm-adjusted 60-mer probes covering the entire yeast genome with virtually no gaps between probes. All ChIP-chip experiments were done in duplicate.

### **NET-seq experiments**

The NET-seq protocol was adapted from (Churchman and Weissman, 2011). Yeast cells were inoculated at  $OD_{600}=0.1$  in 1 L of YNB –HIS supplemented with 2% glucose and grown at 30°C. When cells reached  $OD_{600}$  0.7-0.8, they were harvested by microfiltration. Cells were scraped off the filter and transferred into a 50 mL canonical tube containing liquid nitrogen. Liquid nitrogen was removed and cells were stored at -80°C.

Frozen cells were cryogenically lysed in a mixer mill (Retasch PM100) using the following parameters: 8 cycles at 500 rpm during 3 min with 1 min reverse rotation without breaks between rotations, keeping cells frozen at all time. Frozen cell grindates were stored at -80°C. 1 g of cell grindate was resuspended in 5 mL of ice-cold Lysis buffer (20 mM HEPES pH 7.4, 110 mM KOAc, 0.5% Triton X-100, 0.1% Tween-20, 10 nM MnCl<sub>2</sub>, 1 mM PMSF, 1 mM benzamidine, 10 µg/mL aprotinin, 1 µg/mL leupeptin and 1 µg/mL pepstatin) by pipetting. 660 units of RNase-free DNase I (Promega, MG101) were added to the mix and the sample was incubated 20 min on ice. The lysate was then centrifuged for 10 min at 20,000 g at 4°C. 300 µL of Dynabeads coated with Pan Mouse IgG antibodies (Thermo Fisher Scientific, 11042), pre-coupled to 30 µg of anti-Flag antibody (Sigma-Aldrich, F3165) (as per the ChIP-chip protocol) were added to the supernatant which was incubated 2.5 hr at 4°C on rotator. Beads were washed four times with 10 mL of ice-cold Wash buffer (20 mM HEPES pH 7.4, 110 mM KOAc, 0.5% Triton X-100, 0.1% Tween-20, 1 mM EDTA). After the last wash, beads were transferred into a 1.5 mL microcentrifuge tube and placed on a magnet to discard



any remaining supernatant. To eluate the IP proteins, 120 mg of 3xFlag peptide (Sigma-Aldrich, F4799-4MG) in 600  $\mu$ L of ice-cold Lysis buffer was added to the beads which were incubated 30 min on ice. A second elution was performed and both supernatants were combined. The immunoprecipitated RNA was purified using the QIAgen miRNeasy mini kit (217004) according to manufacturer's protocol with the following modifications: 1.5 volumes of QIAzol lysis reagent were used and two elutions with 50  $\mu$ L nuclease-free water were done and pooled. RNA was quantified on a Nanodrop (ThermoFisher), precipitated with ethanol and resuspended in 11  $\mu$ L of 10 mM Tris-HCl pH 7.0.

To construct a DNA sequencing library from the 3' ends of a nascent RNA sample, 5  $\mu$ L of RNA (1.5  $\mu$ g) in 10 mM Tris-HCl pH 7.0 was denatured for 2 min at 80°C in a thermocycler and placed on ice. DNA adapters (all from (Ingolia et al., 2012); see **Table S7**) were ligated to each RNA sample by adding 14  $\mu$ L ligation reaction mix (8  $\mu$ L 50% PEG M.W. 8,000, 2  $\mu$ L 10X RNL2 buffer, 1  $\mu$ L DNA linker (0.5  $\mu$ g/ $\mu$ L; see **Table S7** for a list of adapters with their bar codes), 3  $\mu$ L nuclease-free H<sub>2</sub>O and 1  $\mu$ L (200 units) T4 RNA ligase 2 (truncated, K227Q; New England Biolabs M0351S) and incubated overnight at 16°C. Tubes were placed on ice and 0.7  $\mu$ L 0.5 M EDTA was added. Then, 20  $\mu$ L of 2X alkaline fragmentation solution (100 mM NaCO<sub>3</sub> pH 9.2, 2 mM EDTA) was added to the ligated RNA samples which were incubated at 95°C for a pre-determined calibrated time (0-50 min). The fragmented RNA was precipitated with isopropanol, washed with 70% ethanol and resuspended in 10  $\mu$ L 10 mM Tris-HCl pH 7.0. The RNA was run on 15% TBE-urea polyacrylamide gel, stained with SYBR gold. Dark Reader DR-45M2 transilluminator (Clare Chemical Research) was used to visualize the RNA and the gel piece corresponding to the ligated oligonucleotide (between 45 nt and 95 nt) was excised. Gel pieces were placed into 0.5 mL RNase-free, non-stick microcentrifuge tubes that had been needle-pierced and placed inside 1.5 mL RNase-free non-stick tubes. Tubes were centrifuged 3 min at 20,000 g at room temperature to force the gel through the needle hole. 200  $\mu$ L RNase-free H<sub>2</sub>O was added, followed by 10 min incubation at 70°C. The gel slurry was vortexed for 30 sec, transferred to a Spin-X® centrifuge tube filter (Costar) and centrifuged 3 min at 20,000 g at room temperature to recover the elution mixture. RNA was ethanol-precipitated and resuspended in 10  $\mu$ L of 10 mM Tris-HCl pH 7.0. RNA samples were then reverse transcribed as follow. 4.6  $\mu$ L of RT

reaction mix (3.28  $\mu\text{L}$  5X first strand buffer, 0.82  $\mu\text{L}$  dNTPs (10 mM) and 0.5  $\mu\text{L}$  RT primer; see **Table S7** for primer sequence) was added to samples, which were denatured 2 min at 80°C and placed on ice. A mix of 0.5  $\mu\text{L}$  (0.01 units) SUPERase In (ThermoFisher AM2694) and 0.82  $\mu\text{L}$  0.1 M DTT was added together with 0.82  $\mu\text{L}$  (170 units) Superscript IV (ThermoFisher 18090050) and samples were incubated 30 min at 48°C. To degrade RNA, 1.8  $\mu\text{L}$  1 M NaOH was added and samples were incubated 20 min at 98°C. Reactions were neutralized by adding 1.8  $\mu\text{L}$  1 M HCl. Samples were run on 10% TBE-urea polyacrylamide gel, bands were excised and DNA was extracted as described above. cDNA samples were precipitated with isopropanol and resuspended in 15  $\mu\text{L}$  of 10 mM Tris-HCl pH 7.5. A mix of (2  $\mu\text{L}$  10X circ buffer, 1  $\mu\text{L}$  1 mM ATP, 1  $\mu\text{L}$  50 mM  $\text{MnCl}_2$  and 1  $\mu\text{L}$  (100 units) CircLigase (Mandel LGN-CL4111K)) was added to cDNA and samples were incubated 1 hr at 60°C to circularize DNA. Samples were incubated 10 min at 80°C to heat-inactivate the enzyme. 5  $\mu\text{L}$  of circularized DNA was added to 79.2  $\mu\text{L}$  of PCR master mix (16.7  $\mu\text{L}$  5X Phusion HF buffer, 0.7  $\mu\text{L}$  10 mM dNTPs, 0.8  $\mu\text{L}$  50  $\mu\text{M}$  reverse PCR primer (see **Table S7** for sequence), 0.4  $\mu\text{L}$  100  $\mu\text{M}$  forward primer (see **Table S7** for sequence), 58.8  $\mu\text{L}$  RNase-free  $\text{H}_2\text{O}$  and 0.8  $\mu\text{L}$  (1.6 units) Phusion polymerase (New England Biolabs M0530S)) and PCR was performed as follows: 1) 98°C, 30 sec; 2) 6, 8, 10 and 12 cycles of 98°C, 10 sec/60°C, 10 sec/ 72°C, 5 sec; 3) 4°C, forever. PCR products were run on 8% TBE gel and stained with SYBR gold. The appropriate number of cycles (6, 8, 10 or 12 cycles) was determined and the PCR was repeated on the remaining of the sample using the selected number of cycle. PCR bands were cut and centrifuged through a needle hole as described above. Crunched gels were soaked in 670  $\mu\text{L}$  DNA soaking buffer (0.3 M NaCl, 10 mM Tris-HCl pH 8.0 and 0.97 mM EDTA) overnight with agitation at room temperature. The gel mixture was transferred to a Spin-X® centrifuge tube filter (Costar) and centrifuged 30 min at 20,000 g at room temperature. Samples were precipitated with isopropanol and resuspended in 7  $\mu\text{L}$  of 10 mM Tris-HCl pH 8.0. Library concentration and size (expected to be in the 175-225 bp range) were determined on a 2100 Bioanalyzer (Agilent). Sequencing was performed on an Illumina HiSeq 2500 platform at the McGill University and Génome Québec Innovation Centre to generate 50 bp singled-end reads (see **Table S4** for sequencing primers).

#### 2.3.9.4 Quantification and statistical analysis

##### ChIP-chip analyses

The ChIP-chip data were normalized using the Limma Loess method, and replicates were combined as described previously (Ren et al., 2000). The data were subjected to one round of smoothing using a Gaussian sliding window with a standard deviation of 100 bp to generate data points in 10-bp intervals (Guillemette et al., 2005).

##### NET-seq data analysis

Reads were aligned to the *S. cerevisiae* (sacCer3) genome using TopHat2 (Kim et al., 2013) with the following parameters: Read mismatches = 3, Read gap length = 2, Read edit distance = 3, Minimum anchor length = 8, Splice mismatches = 1, Minimum intron length = 50, Maximum intron length = 1200, Maximum insertion length = 3, Maximum deletion length = 3, Maximum multihits = 100, Segment mismatches = 3, Minimum coverage intron = 50, Maximum coverage intron = 100000, Minimum segment intron = 50, Maximum segment intron = 500000. The resulting BAM files were converted to BED files using Bamtools (Barnett et al., 2011) and then to bigBed using ucsc\_tools ([http://hgdownload.soe.ucsc.edu/downloads.html#source\\_downloads](http://hgdownload.soe.ucsc.edu/downloads.html#source_downloads)). The reads were trimmed to keep only the first sequenced base using an in-house tool available at <https://github.com/IRCM/bed-tools-j/releases/tag/2>. Coverage for each base pair of the *S. cerevisiae* genome (on each strand) was computed using genomeCoverageBed from BEDTools (version 2.19.1) (Quinlan and Hall, 2010) and normalized to 100 million uniquely mapped reads (as determined by running Flagstat from Samtools (version 1.2) (Li et al., 2009) on the BAM files). BedGraphs were then converted to bigWig (using Wig/BedGraph-to-bigWig from ucsc\_tools) for data visualization on the UCSC Genome Browser (Casper et al., 2018) and for downstream analyses (see below). Replicate samples showed high correlation and were combined by merging BAM files using MergeSamFiles from Picard Tools (version 1.56.0) (<http://broadinstitute.github.io/picard/>) prior to calculation of the coverage and normalization.

## Metagene and Heat map profiles

Average and heat map profiles were done using Versatile aggregate profiler (VAP) (Brunelle et al., 2015; Coulombe et al., 2014). Coordinates for the transcription termination site (TTS) sites were from (Xu et al., 2009) when mapping ChIP-chip data and from (Pelechano et al., 2013) when mapping NET-seq data. Four reference points were used for the termination factors profiles at mRNA-coding genes while two were used for the factors at snoRNAs with a window of 10 bp. Pcf11, Rtt103 and RNAPII average occupancies around the TTS were determined for protein-coding genes that are longer than 500 bp, have an average RNAPII log<sub>2</sub> occupancy greater than 0.5 (expressed genes) and are at least 500 bp away from an expressed downstream neighbouring gene (irrespective of its orientation) (n=625; listed in **Table S2**, column A). RNAPII average occupancies in anchored-away Glc7 (this study and (Schrieck et al., 2014)) around the TTS were determined for protein-coding genes expressed in both WT and Glc7 anchor-away cells that are longer than 500 bp and are at least 500 bp away from an expressed downstream neighbouring gene (irrespective of its orientation) (n=452 (Glc7-anchored-away from this study), n=152 (Glc7-anchored-away from Schrieck et al., 2014); listed in **Table S2**, columns B and C respectively). Nrd1, Sen1 and RNAPII average occupancies were determined for snoRNAs fulfilling the following criteria: 1) The gene is not localized within the intron of another gene; 2) the gene is not contained within a polycistronic cluster of snoRNAs, except the last one; 3) the gene is at least 500 bp away from a tandem and expressed neighbouring gene; 4) the gene does not have a convergent expressed neighboring gene (n=33, listed in **Table S2**, column D). NET-seq data were mapped on all genes with an average signal >0.3 (n=712, listed in **Table S1**, column A) by aligning either on their TSS or TTS. In **Figures 6D** and **S6**, the NET-seq data was normalised to set the average signal for each gene equal in both WT and Y1F cells. This was done to better emphasise local difference in signals (such as pause sites) relative to effects on global occupancy.

## Termination factor occupancy normalization

In **Figures 2C**, **5A** (left panel), **5B** (left panel), **5C** (left panel) and **5D** (left panel), the enrichment of various termination factors is shown in CTD, Nrd1 and Sen1 mutants. To

remove biases due to differences in RNAPII occupancy in the various mutants, the enrichment of the termination factors were normalized to the level of RNAPII at the TTS. The difference between the  $\log_2$  occupancy ratio of the WT and mutant RNAPII at the TTS was applied to the termination factor occupancy. Not-normalized data for these experiments are also shown in **Figures S2 and S5**.

### **Readthrough index calculation from ChIP-chip data**

To calculate a readthrough index (RTI) for each gene from ChIP-chip data, the maximum RNAPII occupancy value of each gene was calculated and the downstream coordinate  $X$  corresponding to 25% of this value was determined. The distance between  $X$  obtained in WT and  $X$  obtained in the tested mutant was computed and used as the RTI. Note that the 25% value used here is arbitrary and that the analyses led to similar outputs using various percentages ranging from 10% to 40% (not shown). We applied a threshold of 100 bp to these RTI based on the distribution of the WT control. The genes included in these analyses were selected based on the following criteria. For snoRNAs, the criteria were the same as those described in the “Metagene profiles” section above (n=33, listed in **Table S2**, column D). For CUTs, we have excluded those that are intragenic to an expressed gene on the same strand. We also only considered CUTs that were at least 500 bp away from an expressed neighbouring tandem gene. Finally, we excluded CUTs that have a convergent expressed neighboring gene (n=58, listed in **Table S2**, column E). For protein-coding genes, we used expressed genes that are longer than 500 bp (n=625, listed in **Table S2**, column A). In **Figure 2B**, we restricted the display to expressed genes that are longer than 500 bp from which we removed those that have a downstream tandem and expressed neighbour less than 500 bp away (n=300, listed in **Table S2**, column F). This filter has no impact on the traces for any CTD mutant except for S5A. For that mutant, this filtering allows for removing artifacts emanating from the downstream neighbour, due to the strong promoter-escape defect of the S5A polymerase.

After applying the indicated threshold, these analyses identified 207 protein-coding genes, 31 snoRNAs and 34 CUTs to be dependent on Tyr1 for termination (listed in **Table S2**

columns G, H and I, respectively). We also identified 183 protein-coding genes and 24 snoRNAs to be dependent on Glc7 for termination based on our Glc7 anchor-away experiment (listed in **Table S2** columns J and K, respectively). Finally, we identified 23 snoRNAs that are dependent on Nrd1 for termination based on our Nrd1 anchor-away experiment (listed in **Table S2**, column L).

### **Termination defect index calculation from NET-seq data**

The TDI from the NET-seq data was defined as the read count 250 bp downstream of the TTS, divided by the read count 250 bp upstream of the TTS. The threshold was determined using the index obtained comparing both replicates from the WT. We used a threshold that gives a FDR of 5% based on the WT replicates. The genes included in these analyses were selected using the same criteria irrespective of them being a protein-coding gene, a snoRNA or a CUT. Essentially, any expressed gene (average reads per million (RPM) $>0.3$ ) that is at least 250 bp away from the neighbouring expressed gene was used. For protein-coding genes, we also required the gene being longer than 500 bp. The 3' end coordinate of the snoRNAs were from (Roy et al., 2016) (Protein-coding genes, n=712; snoRNAs, n=49; CUTs, n=99; all listed in **Table S1**, columns A, D and F, respectively).

After applying the indicated threshold, these analyses identified 67 protein-coding genes, 46 snoRNAs and 73 CUTs to be dependent on Tyr1 for termination (listed in **Table S1** columns B, E and G, respectively). Tyr1-independent protein-coding genes (n=645) are listed in **Table S1** column C.

### **Identification of pause sites genome-wide**

Pause sites were identified genome-wide using the NET-seq data from WT cells using scripts generously provided by Stirling Churchman and adapted locally. Briefly, pause sites are defined as coordinates where the NET-seq signal is at least three standard deviations above the mean signal of the surrounding 200 bp excluding bp contained within another pause site. Pause sites were also required to have a minimum of four reads. Our implementation of this

algorithm is available at <https://github.com/IRCM/rna-polymerase-pauses-j/releases/tag/2> and coordinates of pause sites identified from our WT NET-seq data are available in **Table S3** (tab1, n=23,045). For the heat maps shown in **Figures 6D** and **S6B**, only pause sites contained within expressed genes were used. Also, when multiple pause sites were identified within a window of 20bp, only the one showing the maximum score was kept **Table S3** (tab2, n=10,661).

### **Data and software availability**

ChIP-chip and NET-seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE115028.

## 2.3.10 Supplemental Information

### 2.3.10.1 Figures

Figure S2.1

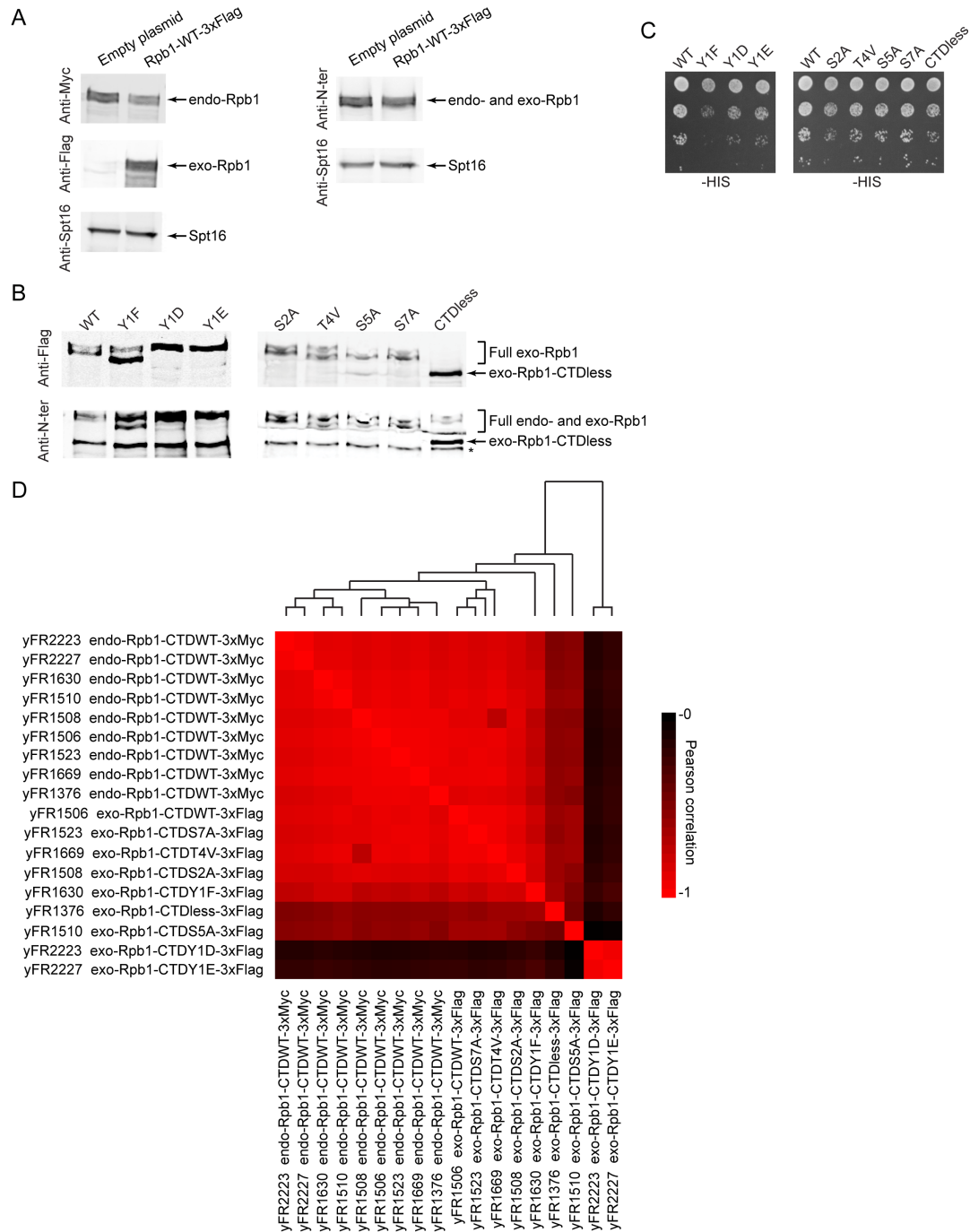




Figure S2.2

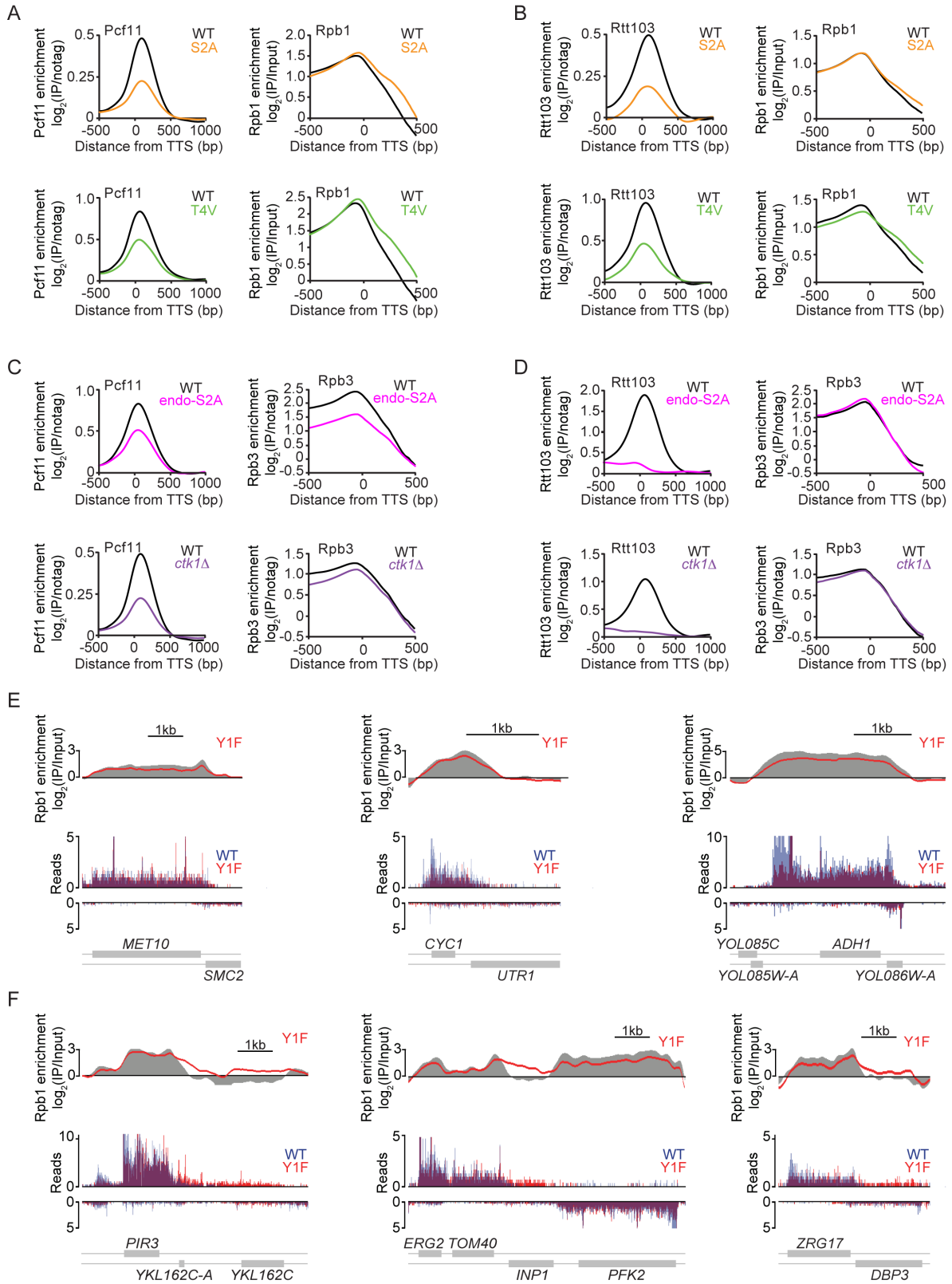


Figure S2.3

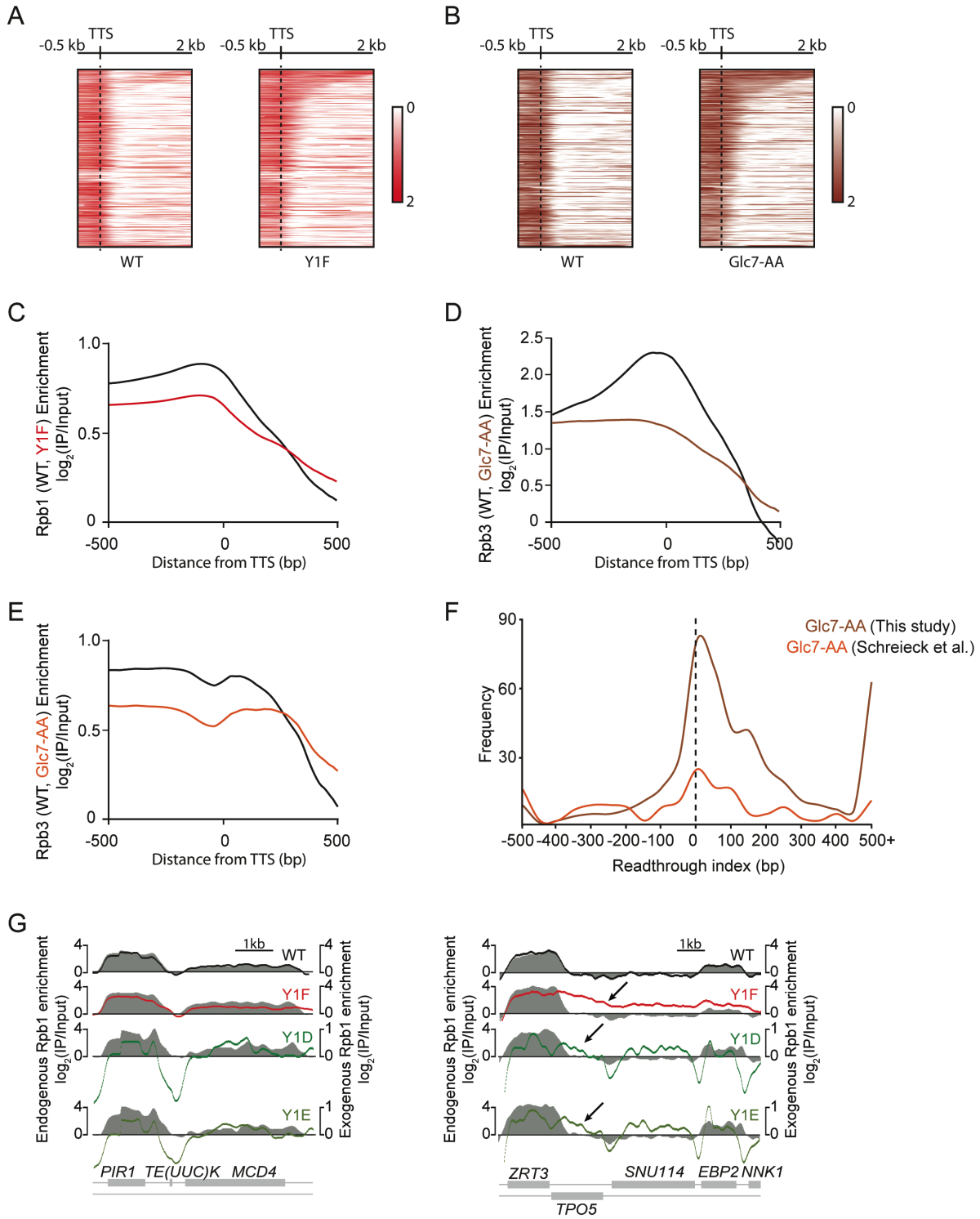


Figure S2.4

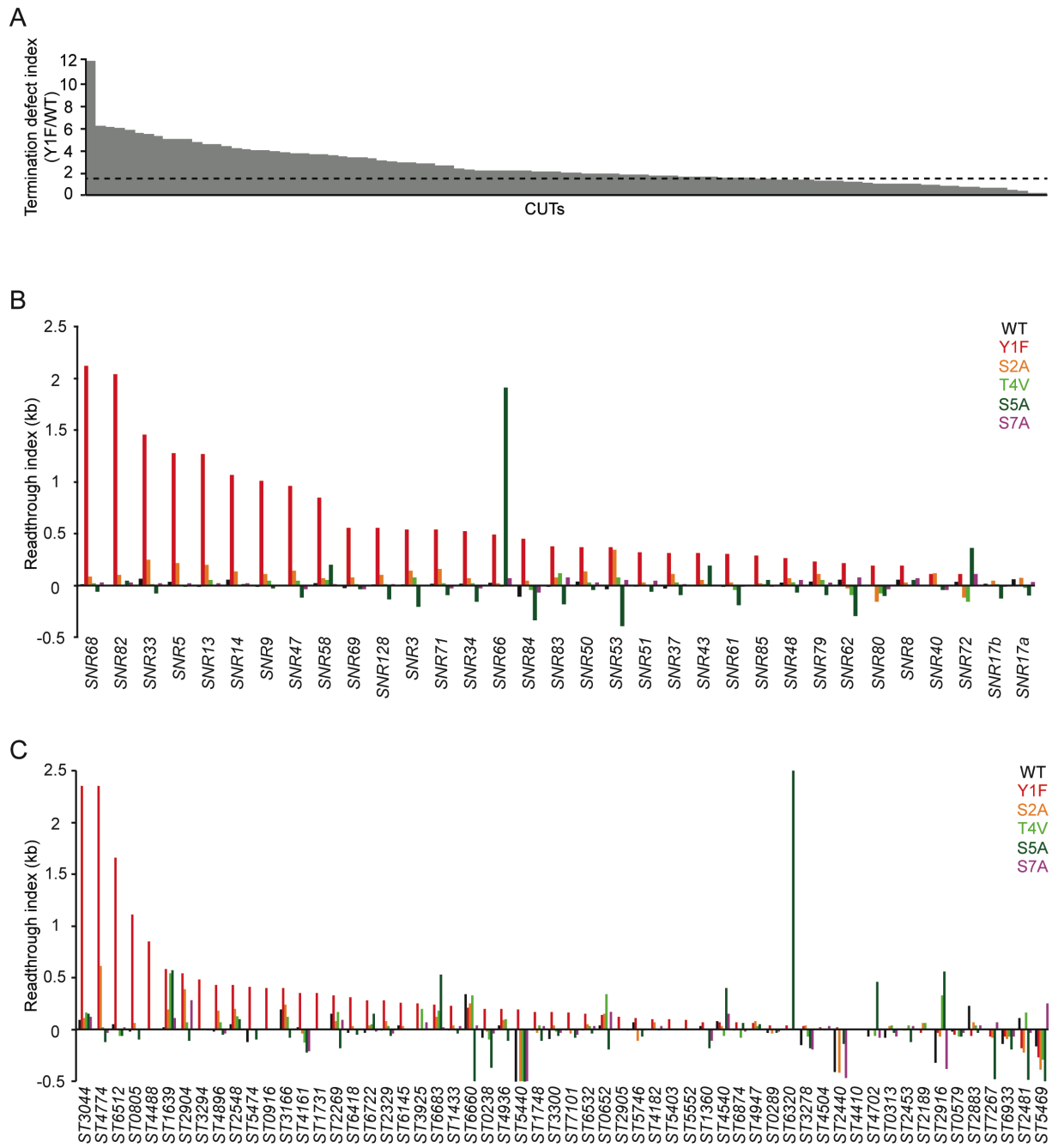


Figure S2.5

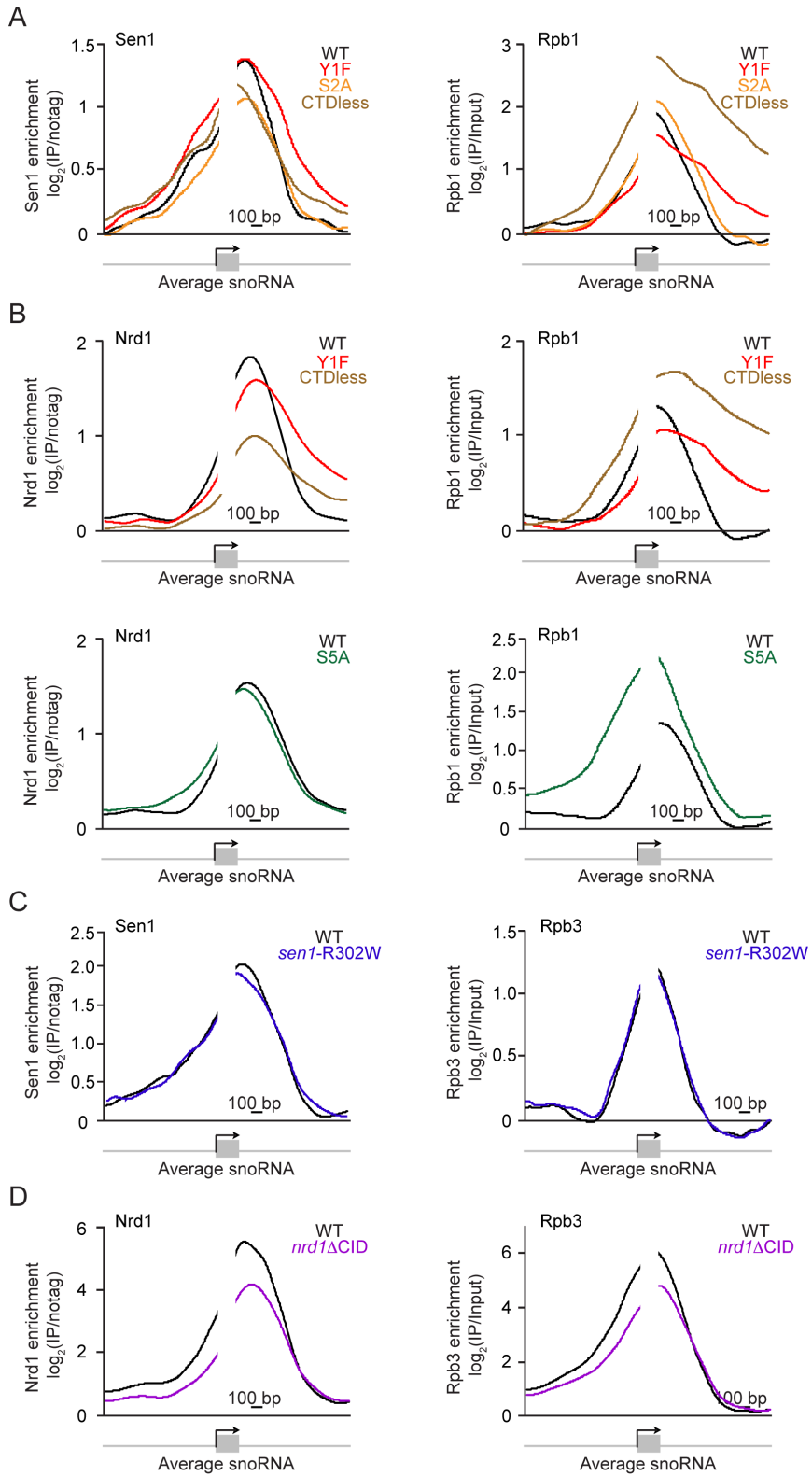
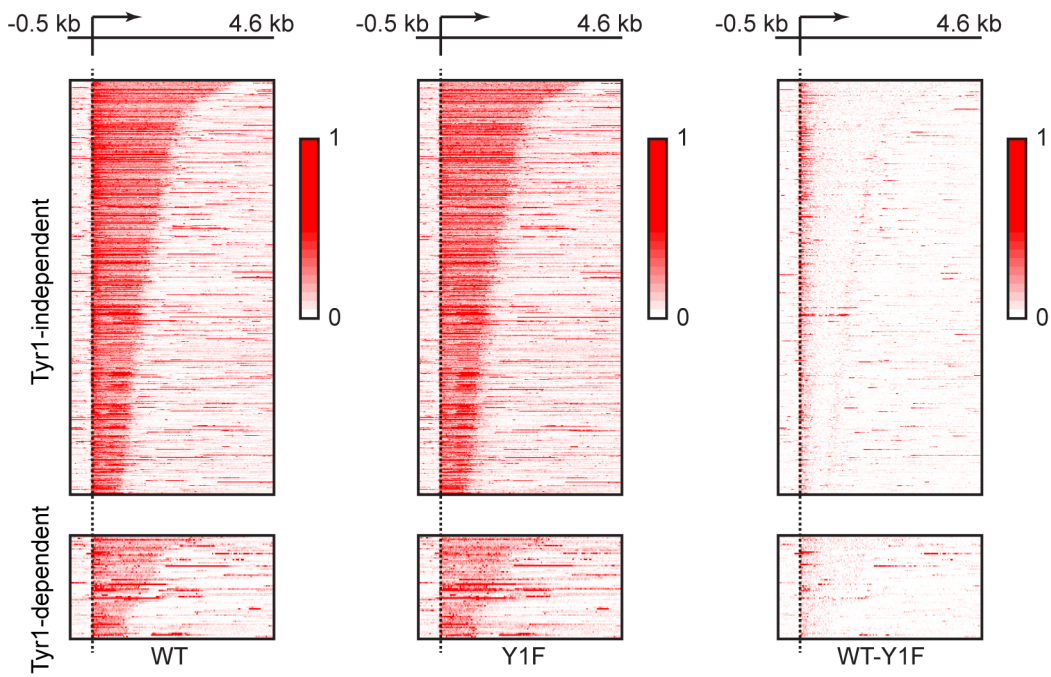


Figure S2.6

A



B

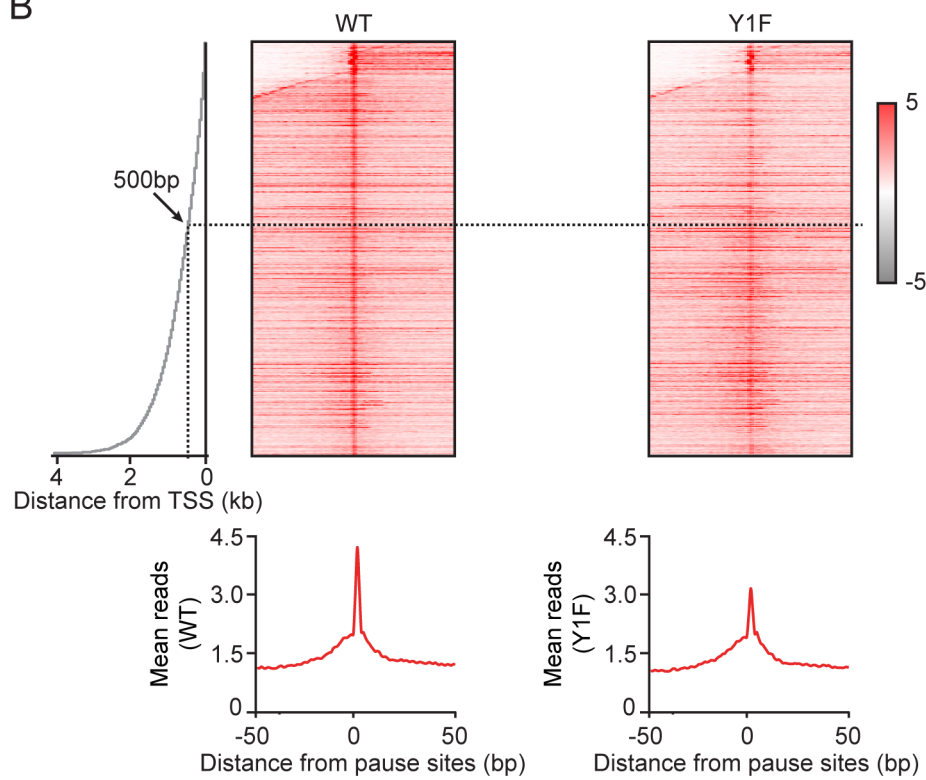
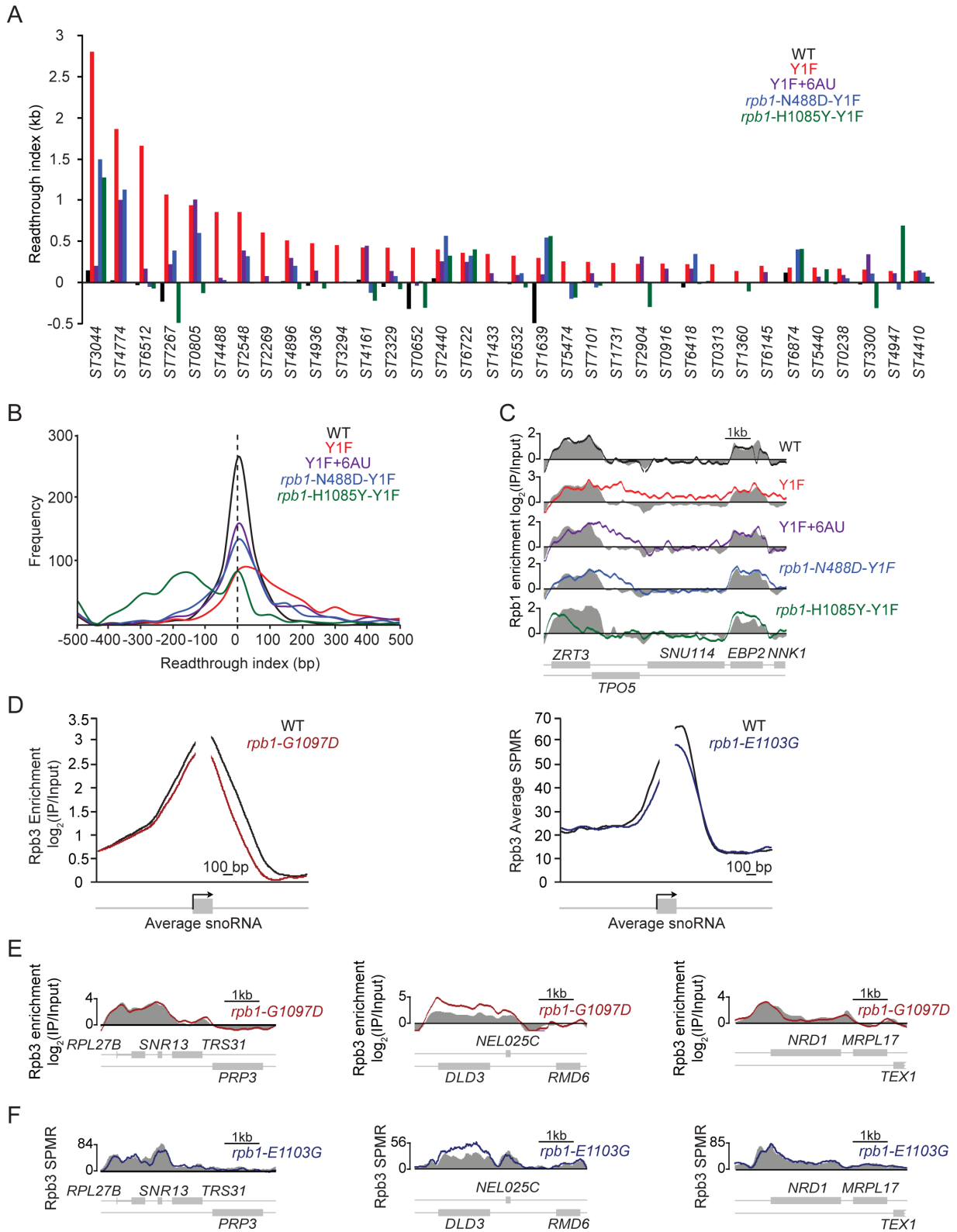


Figure S2.7



### 2.3.10.2 Figure legends

#### Figure S2.1 Related to Figure 2.1.

Related to Figure 1. A) Immunoblot of cell extracts from strains expressing the endogenous RPB1-WT-3xMyc and either an exogenous RPB1-WT-3xFlag or an empty plasmid. Membranes were probed with anti-Myc, anti-Flag, anti-Rpb1-N-terminal (anti-N-ter) and anti-Spt16 as indicated. B) Immunoblots of cell extracts from strains expressing the endogenous RPB1-WT-3xMyc and different exogenous *rpb1*-CTD-mutant-3xFlag. The top panel shows a membrane probed with anti-Flag and the bottom panel shows the same membrane probed with an antibody recognizing the N-terminal region of Rpb1. The asterisk indicates a non-specific band. C) Spotting assay for strains expressing the endogenous RPB1-WT-3xMyc and different exogenous *rpb1*-CTD-mutant-3xFlag. D) Heat map representation of Pearson correlations between the ChIP-chip profiles of various endogenous RPB1 and exogenous *rpb1* alleles after hierarchical clustering.

#### Figure S2.2 Related to Figure 2.2.

A) Non-normalized average occupancy of Pcf11 (left) and exogenous Rpb1 (right) in WT (black), S2A (orange) and T4V (green) cells around the TTS sites of expressed mRNA-coding genes (n=625, see STAR Methods) after nuclear depletion of the endogenous Rpb1 by anchor-away. B) Non-normalized average occupancy of Rtt103 (left) and exogenous Rpb1 (right) in WT (black), S2A (orange) and T4V (green) cells around the TTS sites of expressed mRNA-coding genes (n=625, see STAR Methods) after nuclear depletion of the endogenous Rpb1 by anchor-away. C) Pcf11 (left) and RNAPII (Rpb3; right) average occupancies around the TTS sites of transcriptionally active mRNA-coding genes (n=625, see STAR Methods) in WT (black), endogenous CTD-S2A allele (pink) and *ctk1Δ* (purple) strains. D) Rtt103 (left) and RNAPII (Rpb3; right) average occupancies around the TTS sites of transcriptionally active mRNA-coding genes (n=625, see STAR Methods) in WT (black), endogenous CTD-S2A allele (pink) and *ctk1Δ* (purple) strains. Note that termination defects in the endogenous CTD-S2A and *ctk1Δ* strains are not as strong as in strains where the CAND system is used (as in Figure 2). While we do not have a definite answer as to why that is, we have array-based comparative genomic hybridization (aCGH) data showing that these strains have accumulated

chromosomal abnormalities (not shown). It is therefore possible that these strains have accumulated suppressor mutations alleviating their termination defects. E) Top: The enrichment of endogenous WT (grey shadow) and Y1F (red trace) RNAPII (Rpb1), as determined by ChIP-chip is shown over the MET10 (left), CYC1 (middle) and ADH1 (right) genes. Bottom: For the same genes, the NET-seq signal from WT (blue) and Y1F (red) cells are shown. The overlap is shown in purple. F) Top: The enrichment of endogenous WT (grey shadow) and Y1F (red trace) RNAPII (Rpb1), as determined by ChIP-chip is shown over the PIR3 (left), TOM40 (middle) and ZRG17 (right) genes. Bottom: For the same genes, the NET-seq signal from WT (blue) and Y1F (red) cells are shown. The overlap is shown in purple.

**Figure S2.3 Related to Figure 2.2.**

A) Left: Heat map representation of the WT and Y1F RNAPII occupancy, as determined by ChIP-chip, over expressed protein-coding genes (n=625, see STAR Methods). Right: Heat map representation of RNAPII occupancy, as determined by ChIP-chip, in WT and Glc7-AA cells after 90 minutes of rapamycin treatment, over expressed protein-coding genes (n=452, see STAR Methods). Genes are sorted based on decreasing readthrough index. B) WT (black traces) and Y1F (red traces) RNAPII (Rpb1) average occupancy around the TTS sites of expressed protein-coding genes (n=625, see STAR Methods). C) RNAPII (Rpb3) average occupancy in WT (black traces) and Glc7-AA from this study (brown traces) cells around the TTS sites of expressed protein-coding genes (n=452, see STAR Methods). D) RNAPII (Rpb3) average occupancy in WT (black traces) and Glc7-AA from (Schrieck et al., 2014) (brown traces) cells around the TTS sites of expressed protein-coding genes (n=152, see STAR Methods). E) Distribution of the readthrough index at expressed protein-coding genes in Glc7-AA from this study (brown; n=452, see STAR Methods) and from (Schrieck et al., 2014) (orange; n=152, see STAR Methods). F) The enrichment of endogenous WT RNAPII (grey shadow) and RNAPII carrying the Y1F, Y1D or Y1E CTD mutations (colored traces) is shown over the PIR1 (left) and ZRT3 (right) genes. Arrows indicate termination defects. Note that the scale of the y axis for the Y1D and Y1E mutants was adjusted to accommodate the low RNAPII occupancy level in these mutants.



**Figure S2.4 Related to Figure 2.3.**

A) Termination defect index, determined from the NET-seq profile of the Y1F mutant at CUT genes (n=99, see STAR Methods). The dashed line represents the threshold used (FDR=5%). B-C) Readthrough index, based on the ChIP-chip experiments, for all CTD mutants at snoRNA (B) (n=33, see STAR Methods) and CUT (C) genes (n=58, see STAR Methods).

**Figure S2.5 Related to Figure 2.5.**

A) Non-normalized Sen1 (left) and RNAPII (Rpb1, right) average occupancy over snoRNA loci (n=33, see STAR Methods) in the indicated CTD mutants after nuclear depletion of the endogenous Rpb1. B) Non-normalized Ndr1 (left) and RNAPII (Rpb1, right) average occupancy over snoRNA loci (n=33, see STAR Methods) in the indicated CTD mutants after nuclear depletion of the endogenous Rpb1. C) Non-normalized Sen1 (left) and RNAPII (Rpb3, right) average occupancy over snoRNA loci (n=33, see STAR Methods) in WT (black) and *sen1*-R302W (blue) cells. D) Non-normalized Nrd1 (left) and RNAPII (Rpb3, right) average occupancy over snoRNA loci (n=33, see STAR Methods) in WT (black) and *nrd1*ΔCID (purple) cells.

**Figure S2.6 Related to Figure 2.6.**

A) Heat map representation of the normalized NET-seq signal from WT (left) and Y1F (middle) cells, together with their difference (WT-Y1F, right), over all expressed protein-coding genes not requiring (top; n=645, see STAR Methods) or requiring (bottom; n=67, see STAR Methods) Tyr1 for termination, sorted by decreasing gene length. B) Heat map representation of the normalized NET-seq signal from WT (left) and Y1F (right) cells around pause sites ordered by their distance from the TSS (shown on the left). The average signal is also shown as a trace at the bottom.

**Figure S2.7 Related to Figure 2.7.**

A) Readthrough index at CUTs dependent on Tyr1 for efficient termination (n=34, see STAR Methods) in WT, Y1F, Y1F treated with 6AU, *rpb1*-N488D-Y1F and *rpb1*-H1085Y-Y1F cells. B) Distribution of RTI in WT, Y1F, Y1F treated with 6AU, *rpb1*-N488D-Y1F and *rpb1*-

H1085Y-Y1F cells for expressed protein-coding genes (n=625, see STAR Methods). C) The enrichment of endogenous WT (grey shadow) and various RNAPII mutants treated or not with 6AU (colored traces) is shown over *ZRT3* gene. D) Average occupancy of RNAPII (Rpb3) in WT (black), *rpb1*-G1097D (brown, left) and *rpb1*-E1103G (blue, right) cells at snoRNAs (n=33, see STAR Methods). E) The enrichment of RNAPII (Rpb3) in WT (grey shadow) and *rpb1*-G1097D cells (brown) is shown over the *SNR13*, *NEL025C* and *NRD1* loci. F) The enrichment of RNAPII (Rpb3) in WT (grey shadow) and *rpb1*-E1103G cells (blue), as determined by CHIP-seq in (Soares et al., 2017), is shown over the *SNR13*, *NEL025C* and *NRD1* loci. SPMR, sequence tags per million reads.

### 2.3.10.3 Tables

#### **Table S2.1 Gene lists from NET-seq experiments.**

Related to STAR Methods. Available online.

#### **Table S2.2 Gene lists from CHIP-chip experiments.**

Related to STAR Methods. Available online.

#### **Table S2.3 Pause sites identified in this study.**

Related to STAR Methods. Available online.

**Table S2.4 Yeast strains used in this study.**

Related to STAR Methods.

Strain	Alias	Genotype	Source
yFR116	Z1256	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+	A kind gift of R.A. Young
yFR117		MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0	ATCC
yFR1308		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3	This study
yFR3123		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pRS313 (CEN, HIS3)]	This study
yFR1376		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR467 (CEN, HIS3, rpb1-CTDless-3xFLAG)]	This study
yFR1506		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR482 (CEN, HIS3, RPB1-CTDWT-3xFLAG)]	This study
yFR1508		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR486 (CEN, HIS3, rpb1-CTDS2A-3xFLAG)]	This study
yFR1510		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR490 (CEN, HIS3, rpb1-CTDS5A-3xFLAG)]	This study
yFR1523		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR494 (CEN, HIS3, rpb1-CTDS7A-3xFLAG)]	This study
yFR1630		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR511 (CEN, HIS3, rpb1-CTDY1F-3xFLAG)]	This study
yFR1669		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR508 (CEN, HIS3, rpb1-CTDT4V-3xFLAG)]	This study
yFR2223		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR574 (CEN, HIS3, rpb1-CTDY1D-3xFLAG)]	This study
yFR2227		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR576 (CEN, HIS3, rpb1-CTDY1E-3xFLAG)]	This study
yFR2769		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR683 (CEN, HIS3, rpb1-H1085Y-CTDWT-3xFLAG)]	This study
yFR2770		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR684 (CEN, HIS3, rpb1-H1085Y-CTDY1F-3xFLAG)]	This study
yFR2959		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR686 (CEN, HIS3, rpb1-H1085Y-CTDWT-3xFLAG)]	This study
yFR2960		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, RPB1-3xMYC::URA3, [pFR687 (CEN, HIS3, rpb1-N488D-CTDY1F-3xFLAG)]	This study
yFR1240	CKY1016	MATa, lys2-128 $\hat{c}$ , ura3-52, his3Δ200, leu2Δ0 or Δ1, trp1Δ63, met15Δ0, rpb1Δ::CLONATMX, RPB3-3XFLAG::KanMX, [pRS315 (CEN, LEU2, RPB1)]	A kind gift of C.D. Kaplan

yFR1252	CKY1028	MAT $\alpha$ , lys2-128 $\delta$ , ura3-52, his3 $\Delta$ 200, leu2 $\Delta$ 0 or $\Delta$ 1, trp1 $\Delta$ 63, met15 $\Delta$ 0, rpb1 $\Delta$ ::CLONATMX, RPB3-3XFLAG::KanMX, [pRS315 (CEN, LEU2, rpb1-G1097D)]	A kind gift of C.D. Kaplan
yFR1321	HHY168	MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1	Euroscarf
yFR1324	HHY170	MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6	Euroscarf
yFR1495		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, [pFR467 (CEN, HIS3, rpb1-CTDless-3xFLAG)]	This study
yFR1499		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, [pFR482 (CEN, HIS3, RPB1-CTDWT-3xFLAG)]	Jeronimo et al., 2014
yFR1501		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, [pFR486 (CEN, HIS3, rpb1-CTDS2A-3xFLAG)]	Jeronimo et al., 2014
yFR1503		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, [pFR490 (CEN, HIS3, rpb1-CTDS5A-3xFLAG)]	Jeronimo et al., 2014
yFR1632		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, [pFR511 (CEN, HIS3, rpb1-CTDY1F-3xFLAG)]	This study
yFR1667		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, [pFR508 (CEN, HIS3, rpb1-CTDT4V-3xFLAG)]	This study
yFR1824		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, PCF11-3xMYC::URA3, [pFR482 (CEN, HIS3, RPB1-CTDWT-3xFLAG)]	This study
yFR1828		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, PCF11-3xMYC::URA3, [pFR486 (CEN, HIS3, rpb1-CTDS2A-3xFLAG)]	This study
yFR1834		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, PCF11-3xMYC::URA3, [pFR508 (CEN, HIS3, rpb1-CTDT4V-3xFLAG)]	This study
yFR1848		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, RTT103-3xMYC::URA3, [pFR482 (CEN, HIS3, RPB1-CTDWT-3xFLAG)]	This study
yFR1852		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, RTT103-3xMYC::URA3, [pFR486 (CEN, HIS3, rpb1-CTDS2A-3xFLAG)]	This study
yFR1858		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, RTT103-3xMYC::URA3, [pFR508 (CEN, HIS3, rpb1-CTDT4V-3xFLAG)]	This study
yFR1838		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, NRD1-3xMYC::URA3, [pFR467 (CEN, HIS3, rpb1-CTDless-3xFLAG)]	This study
yFR1939		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, NRD1-3xMYC::URA3, [pFR482 (CEN, HIS3, RPB1-CTDWT-3xFLAG)]	This study
yFR1940		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, NRD1-3xMYC::URA3, [pFR511 (CEN, HIS3, rpb1-CTDY1F-3xFLAG)]	This study
yFR1942		MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, NRD1-3xMYC::URA3, [pFR490 (CEN, HIS3, rpb1-CTDS5A-3xFLAG)]	This study

yFR2298	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, SEN1-3xMYC::URA3, [pFR467 (CEN, HIS3, rpb1-CTDless-3xFLAG)]	This study
yFR2292	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, SEN1-3xMYC::URA3, [pFR482 (CEN, HIS3, RPB1-CTDWT-3xFLAG)]	This study
yFR2293	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, SEN1-3xMYC::URA3, [pFR511 (CEN, HIS3, rpb1-CTDY1F-3xFLAG)]	This study
yFR2294	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, SEN1-3xMYC::URA3, [pFR486 (CEN, HIS3, rpb1-CTDS2A-3xFLAG)]	This study
yFR2315	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR482 (CEN, HIS3, RPB1-CTDWT-3xFLAG)]	This study
yFR2316	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::kanMX6, rrp6Δ::URA3, [pFR511 (CEN, HIS3, rpb1-CTDY1F-3xFLAG)]	This study
yFR2317	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR486 (CEN, HIS3, rpb1-CTDS2A-3xFLAG)]	This study
yFR2318	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR490 (CEN, HIS3, rpb1-CTDS5A-3xFLAG)]	This study
yFR2319	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR494 (CEN, HIS3, rpb1-CTDS7A-3xFLAG)]	This study
yFR2320	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR508 (CEN, HIS3, rpb1-CTDT4V-3xFLAG)]	This study
yFR2321	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR467 (CEN, HIS3, rpb1-CTDless-3xFLAG)]	This study
yFR2683	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR574 (CEN, HIS3, rpb1-CTDY1D-3xFLAG)]	This study
yFR2684	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR576 (CEN, HIS3, rpb1-CTDY1E-3xFLAG)]	This study
yFR3141	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR728 (CEN, HIS3, rpb1-D481A-CTDY1D-3xFLAG)]	This study
yFR3142	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, RPB1-FRB::KanMX6, rrp6Δ::URA3, [pFR729 (CEN, HIS3, rpb1-D481A-CTDY1E-3xFLAG)]	This study
yFR1894	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6	Schrieck et al., 2014
yFR2984	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, [pFR2 (CEN, HIS3)]	This study
yFR2985	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, NRD1-FRB::KanMX6, [pFR2 (CEN, HIS3)]	This study
yFR551	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, ctk1D::KanMX4	Open Biosystems
yFR1773	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0	Cassart et al., 2012
yFR1774	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, RPB1::rpb1-CTDS2A-KanMX6	Cassart et al., 2012
yFR1968	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, PCF11-3xMYC::URA3	This study
yFR1969	MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, ctk1Δ::KanMX4, PCF11-3xMYC::URA3	This study

yFR1983		MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, RTT103-3xMYC::URA3	This study
yFR1970		MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, ctk1Δ::KanMX4, RTT103-3xMYC::URA3	This study
yFR1917		MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, PCF11-3xMYC::URA3	This study
yFR1918		MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, RPB1::rpb1-CTDS2A-kanMX6, PCF11-3xMYC::URA3	This study
yFR2082		MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, RTT103-3xMYC::URA3	This study
yFR2083		MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, RPB1::rpb1-CTDS2A-kanMX6, RTT103-3xMYC::URA3	This study
yFR2194	DLY671	MATa, ade2-1, trp1Δ, can1-100, leu2-3,112, his3-11,5, ura3-1	Tudek et al., 2014
yFR2195	DLY2329	MATa, ade2-1, trp1Δ, can1-100, leu2-3,112, his3-11,5, ura3-1, nrd1DCID (D6-150)	Tudek et al., 2014
yFR2208		MATa, ade2-1, trp1Δ, can1-100, leu2-3,112, his3-11,5, ura3-1, NRD1-3xMYC::URA3	This study
yFR2209		MATa, ade2-1, trp1Δ, can1-100, leu2-3,112, his3-11,5, ura3-1, nrd1DCID(D6-150)-3xMYC::URA3	This study
yFR2527		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, SEN1::sen1-R302W	This study
yFR2564		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, SEN1-3xMYC::URA3	This study
yFR2565		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, sen1-R302W-3xMYC::URA3	This study

**Table S2.5 Plasmids used in this study.**

Related to STAR Methods.

<b>Plasmid</b>	<b>Description</b>	<b>Source</b>
pFA6a-FRB-kanMX6	FRB C-terminal tagging	Euroscarf
p3Flag-KanMX	3xFlag C-terminal tagging	Gelbart et al., 2001
pMPY-3xMyc	3xMyc C-terminal tagging	Schneider et al., 1995
pRS313	CEN, HIS3	Sikorski and Hieter, 1989
pFR467	CEN, HIS3, rpb1-CTDless-3xFlag	This study
pFR482	CEN, HIS3, RPB1-CTDWT-3xFlag	Jeronimo et al., 2014
pFR486	CEN, HIS3, rpb1-CTDS2A-3xFlag	Jeronimo et al., 2014
pFR490	CEN, HIS3, rpb1-CTDS5A-3xFlag	Jeronimo et al., 2014
pFR494	CEN, HIS3, rpb1-CTDS7A-3xFlag	Jeronimo et al., 2014
pFR511	CEN, HIS3, rpb1-CTDY1F-3xFlag	This study
pFR508	CEN, HIS3, rpb1-CTDT4V-3xFlag	This study
pFR574	CEN, HIS3, rpb1-CTDY1D-3xFlag	This study
pFR576	CEN, HIS3, rpb1-CTDY1E-3xFlag	This study
pFR683	CEN, HIS3, rpb1-H1085Y-CTDWT-3xFlag	This study
pFR684	CEN, HIS3, rpb1-H1085Y-CTDY1F-3xFlag	This study
pFR686	CEN, HIS3, rpb1-N488D-CTDWT-3xFlag	This study
pFR687	CEN, HIS3, rpb1-N488D-CTDY1F-3xFlag	This study
pML104	2 $\mu$ , URA3, Cas9	Laughery et al., 2015
pFR622	2 $\mu$ , URA3, Cas9, gRNA targeting SEN1 around R302	This study
pFR728	CEN, HIS3, rpb1-D481A-CTDY1D-3xFlag	This study
pFR729	CEN, HIS3, rpb1-D481A-CTDY1E-3xFlag	This study

**Table S2.6 Sequence of each CTD mutant used in this study.**

Related to STAR Methods.

Name	DNA sequence	Protein sequence
CTDWT	1 TTC TCC CCG ACA AGT CCA ACT 2 TAT TCC CCT ACC TCA CCT GCA 3 TAC TCC CCG ACA AGT CCA TCT 4 TAT TCA CCC ACA AGC CCA TCT 5 TAT TCT CCA ACA TCA CCA TCT 6 TAC TCT CCT ACA AGC CCA TCA 7 TAC TCT CCC ACC TCA CCA AGC 8 TAC TCC CCG ACA TCT CCA TCA 9 TAT TCA CCT ACT TCT CCA TCC 10 TAC TCC CCT ACT TCC CCT AGT 11 TAT TCC CCA ACT TCT CCA TCA 12 TAT TCT CCT ACA TCT CCT AGT 13 TAT TCA CCA ACA TCT CCG AGT 14 TAT AGT CCA ACA AGT CCA TCT 15 TAC TCA CCA ACT TCT CCA TCA 16 TAT TCC CCT ACG TCT CCG TCT 17 TAC TCT CCT ACT TCT CCT GCT 18 TAT AGC CCC ACT AGC CCC AGT 19 TAC TCA CCT ACG TCA CCA TCC 20 TAT AGC CCG ACT TCT CCA AGC 21 TAT AGC CCA ACT AGC CCA AGC 22 TAT TCT CCT ACC TCA CCA AAC 23 TAT TCC CCA ACT AGT CCT AGT 24 TAT TCT CCA ACT TCA CCC GGT 25 TAT TCT CCC GGT AGT CCA GCA 26 TAC TCC CCT AAG CAA GAT GAA 27 CAA AAA CAT AAC GAG AAT GAA AAC AGT AGA	1 FSPTSPT 2 YSPTSPA 3 YSPTSPS 4 YSPTSPS 5 YSPTSPS 6 YSPTSPS 7 YSPTSPS 8 YSPTSPS 9 YSPTSPS 10 YSPTSPS 11 YSPTSPS 12 YSPTSPS 13 YSPTSPS 14 YSPTSPS 15 YSPTSPS 16 YSPTSPS 17 YSPTSPA 18 YSPTSPS 19 YSPTSPS 20 YSPTSPS 21 YSPTSPS 22 YSPTSPN 23 YSPTSPS 24 YSPTSPG 25 YSPGSPA 26 YSPKQDE 27 QKHNEENSR
CTDY1F	1 TTT AGC CCT ACA AGC CCA ACT 2 TTC TCT CCT ACC AGT CCT GCT 3 TTC AGT CCT ACA TCT CCA TCT 4 TTT TCT CCT ACA TCC CCT TCA 5 TTT AGT CCA ACA TCA CCA TCC 6 TTT TCA CCA ACG TCT CCC AGT 7 TTC TCA CCA ACT TCC CCA TCA 8 TTC TCA CCT ACG AGC CCA AGC 9 TTT AGC CCC ACC TCA CCC AGT 10 TTT TCC CCA ACC TCT CCC TCT 11 TTT TCA CCA ACC TCT CCT AGT 12 TTT TCA CCT ACA AGT CCA TCA 13 TTC TCT CCT ACT AGC CCT AGT 14 TTT TCC CCT ACC TCT CCA TCT 15 TTT AGT CCT ACC TCC CCA TCT 16 TTC TCA CCA ACG TCA CCT AGC 17 TTT TCA CCA ACG TCT CCA GCG 18 TTT TCA CCA ACT AGC CCT AGT 19 TTC TCA CCA ACA TCA CCA AGT 20 TTT TCT CCT ACA AGC CCC TCA 21 TTT AGC CCG ACA TCT CCA AGC 22 TTT TCC CCC ACC TCT CCT AAC 23 TTC TCC CCA ACA TCA CCA TCA 24 TTC TCC CCG ACT TCT CCA GGC 25 TTC TCT CCA GGT TCA CCC GCT 26 TTC TCA CCT AAA CAA GAC GAA 27 CAG AAA CAT AAT GAA AAT GAA AAC TCT AGA	1 FSPTSPT 2 FSPTSPA 3 FSPTSPS 4 FSPTSPS 5 FSPTSPS 6 FSPTSPS 7 FSPTSPS 8 FSPTSPS 9 FSPTSPS 10 FSPTSPS 11 FSPTSPS 12 FSPTSPS 13 FSPTSPS 14 FSPTSPS 15 FSPTSPS 16 FSPTSPS 17 FSPTSPA 18 FSPTSPS 19 FSPTSPS 20 FSPTSPS 21 FSPTSPS 22 FSPTSPN 23 FSPTSPS 24 FSPTSPG 25 FSPGSPA 26 FSPKQDE 27 QKHNEENSR



CTDY1D	1 TTT TCC CCA ACC TCT CCC ACC 2 GAT TCT CCT ACA TCT CCA GCT 3 GAT AGC CCC ACA TCA CCC TCC 4 GAT AGT CCT ACT AGT CCG TCC 5 GAT AGC CCA ACG AGC CCA TCA 6 GAT AGC CCG ACT TCA CCT AGT 7 GAT TCT CCA ACT TCC CCT TCT 8 GAT TCT CCA ACA TCA CCT TCA 9 GAT TCC CCA ACT AGT CCT TCT 10 GAT AGC CCA ACA TCT CCC TCA 11 GAT TCA CCA ACT TCC CCG TCT 12 GAC TCC CCG ACA AGC CCG TCT 13 GAT TCT CCA ACG TCA CCA TCT 14 GAT TCA CCT ACT TCA CCA TCA 15 GAT TCC CCA ACA TCC CCA TCC 16 GAT TCA CCA ACT AGC CCA TCT 17 GAT TCT CCA ACG TCC CCC GCA 18 GAT AGT CCG ACA AGT CCG TCT 19 GAT TCT CCA ACA TCA CCT TCT 20 GAT TCT CCT ACT TCT CCT TCA 21 GAT TCT CCA ACC AGC CCA TCT 22 GAC TCA CCA ACC TCC CCT AAC 23 GAC TCA CCG ACT TCA CCA TCT 24 GAT TCT CCA ACG TCC CCA GGC 25 GAC AGC CCA GGA AGT CCT GCA 26 GAT AGT CCG AAG CAA GAT GAG 27 CAA AAG CAT AAT GAA AAT GAA AAT TCT AGA	1 FSPTSPT 2 DSPTSPA 3 DSPTSPS 4 DSPTSPS 5 DSPTSPS 6 DSPTSPS 7 DSPTSPS 8 DSPTSPS 9 DSPTSPS 10 DSPTSPS 11 DSPTSPS 12 DSPTSPS 13 DSPTSPS 14 DSPTSPS 15 DSPTSPS 16 DSPTSPS 17 DSPTSPA 18 DSPTSPS 19 DSPTSPS 20 DSPTSPS 21 DSPTSPS 22 DSPTSPN 23 DSPTSPS 24 DSPTSPG 25 DSPGSPA 26 DSPKQDE 27 QKHNEENSR
CTDY1E	1 TTC AGC CCT ACA TCA CCA ACC 2 GAG TCT CCT ACT TCA CCT GCA 3 GAA AGC CCA ACT TCT CCT TCC 4 GAG TCT CCT ACG AGT CCA AGT 5 GAG TCA CCT ACC TCT CCA TCT 6 GAG TCC CCG ACA TCT CCT AGC 7 GAA AGC CCA ACT AGT CCT TCC 8 GAA AGT CCT ACA TCA CCT TCT 9 GAA TCT CCC ACT TCT CCA TCT 10 GAA AGT CCA ACA TCC CCA AGT 11 GAG TCC CCA ACT TCC CCA TCC 12 GAG TCC CCA ACC TCA CCC TCC 13 GAA TCC CCT ACC AGT CCA TCT 14 GAA AGT CCT ACT AGT CCC AGC 15 GAA AGT CCT ACT TCT CCA TCC 16 GAA AGT CCA ACA AGC CCA TCT 17 GAA TCT CCA ACA AGT CCT GCC 18 GAG TCT CCC ACC AGT CCT TCC 19 GAA TCT CCC ACT AGT CCC AGT 20 GAG TCT CCA ACA AGC CCT AGT 21 GAG TCC CCA ACT TCA CCC AGT 22 GAG AGC CCA ACT TCT CCG AAT 23 GAA TCA CCG ACA AGC CCA AGC 24 GAG AGT CCT ACT TCC CCT GGT 25 GAA TCT CCA GGA AGT CCT GCA 26 GAG TCA CCA AAG CAA GAT GAA 27 CAA AAA CAC AAT GAG AAT GAA AAT TCC AGG	1 FSPTSPT 2 ESPTSPA 3 ESPTSPS 4 ESPTSPS 5 ESPTSPS 6 ESPTSPS 7 ESPTSPS 8 ESPTSPS 9 ESPTSPS 10 ESPTSPS 11 ESPTSPS 12 ESPTSPS 13 ESPTSPS 14 ESPTSPS 15 ESPTSPS 16 ESPTSPS 17 ESPTSPA 18 ESPTSPS 19 ESPTSPS 20 ESPTSPS 21 ESPTSPS 22 ESPTSPN 23 ESPTSPS 24 ESPTSPG 25 ESPGSPA 26 ESPKQDE 27 QKHNEENSR

CTDS2A	1 TTT GCT CCT ACT TCA CCT ACA 2 TAC GCT CCT ACC AGC CCG GCC 3 TAC GCA CCT ACC TCA CCA TCT 4 TAT GCA CCT ACG TCA CCT AGT 5 TAT GCG CCT ACT TCC CCT AGT 6 TAT GCC CCA ACT TCC CCA TCT 7 TAT GCG CCT ACC TCT CCT TCT 8 TAT GCT CCC ACA TCT CCA TCA 9 TAC GCA CCG ACA TCA CCA TCC 10 TAC GCC CCC ACT AGT CCA AGC 11 TAT GCA CCG ACT TCT CCT AGT 12 TAC GCT CCA ACT TCA CCT AGC 13 TAT GCA CCA ACT TCC CCT TCT 14 TAC GCC CCG ACT TCA CCG TCA 15 TAT GCC CCT ACA TCT CCA AGC 16 TAC GCT CCG ACA AGT CCA AGT 17 TAC GCC CCA ACT AGT CCC GCT 18 TAT GCG CCA ACC TCC CCA TCA 19 TAT GCA CCT ACG AGT CCA TCT 20 TAT GCG CCA ACC TCA CCA TCT 21 TAC GCT CCT ACC AGT CCC TCA 22 TAT GCA CCA ACT TCA CCT AAT 23 TAT GCA CCT ACA AGT CCT TCT 24 TAC GCT CCA ACC TCA CCA GGA 25 TAC GCG CCA GGT TCT CCG GCT 26 TAT GCA CCA AAG CAG GAT GAA 27 CAG AAA CAT AAT GAG AAT GAA AAc AGC AGA	1 FAPTSPT 2 YAPTSPTA 3 YAPTSPS 4 YAPTSPS 5 YAPTSPS 6 YAPTSPS 7 YAPTSPS 8 YAPTSPS 9 YAPTSPS 10 YAPTSPS 11 YAPTSPS 12 YAPTSPS 13 YAPTSPS 14 YAPTSPS 15 YAPTSPS 16 YAPTSPS 17 YAPTSPA 18 YAPTSPS 19 YAPTSPS 20 YAPTSPS 21 YAPTSPS 22 YAPTSPN 23 YAPTSPS 24 YAPTSPG 25 YAPGSPA 26 YAPKQDE 27 QKHENENSR
CTDT4V	1 TTT TCT CCA GTT TCT CCT ACT 2 TAC TCA CCT GTT TCT CCT GCT 3 TAT AGT CCT GTA TCT CCG AGT 4 TAT TCA CCA GTA TCT CCG TCA 5 TAC TCT CCA GTA AGT CCA TCT 6 TAC AGC CCT GTC TCC CCT AGC 7 TAT TCC CCG GTG TCA CCA AGT 8 TAC AGT CCA GTT TCT CCA TCC 9 TAT TCT CCA GTT AGC CCA TCT 10 TAC AGT CCT GTC TCA CCA AGT 11 TAC TCC CCC GTG AGT CCG TCT 12 TAT TCC CCA GTG TCC CCT TCC 13 TAT TCT CCA GTA TCA CCA TCT 14 TAC AGT CCA GTT TCA CCC TCA 15 TAT AGC CCG GTC AGT CCT TCT 16 TAC TCT CCA GTT AGT CCT TCT 17 TAT TCA CCT GTG TCC CCA GCA 18 TAT AGC CCA GTC TCC CCT TCT 19 TAC AGT CCT GTC AGT CCG AGT 20 TAT AGT CCC GTT TCA CCT AGC 21 TAC AGC CCC GTA AGC CCA TCT 22 TAT TCT CCC GTC TCA CCT AAC 23 TAC AGC CCA GTA TCT CCT TCA 24 TAC TCT CCT GTA AGC CCA GGT 25 TAC TCC CCT GGT TCC CCA GCT 26 TAC AGC CCA AAA CAA GAT GAA 27 CAA AAA CAT AAT GAA AAT GAA AAT TCA AGG	1 FSPVSPT 2 YSPVSPA 3 YSPVSPS 4 YSPVSPS 5 YSPVSPS 6 YSPVSPS 7 YSPVSPS 8 YSPVSPS 9 YSPVSPS 10 YSPVSPS 11 YSPVSPS 12 YSPVSPS 13 YSPVSPS 14 YSPVSPS 15 YSPVSPS 16 YSPVSPS 17 YSPVSPA 18 YSPVSPS 19 YSPVSPS 20 YSPVSPS 21 YSPVSPS 22 YSPVSPN 23 YSPVSPS 24 YSPVSPG 25 YSPGSPA 26 YSPKQDE 27 QKHENENSR

CTDS5A	1 TTT TCC CCG ACT GCC CCG ACA 2 TAT AGC CCT ACA GCC CCG GCG 3 TAT TCA CCA ACC GCA CCA TCT 4 TAT AGC CCC ACT GCG CCC TCT 5 TAC AGC CCC ACT GCC CCA AGC 6 TAT AGC CCG ACA GCA CCC TCA 7 TAT AGC CCG ACT GCG CCA AGC 8 TAT TCT CCA ACA GCC CCG TCA 8 TAT TCA CCA ACT GCC CCA AGT 10 TAT TCC CCG ACG GCT CCA TCT 11 TAC AGC CCG ACC GCT CCC TCT 12 TAC TCT CCT ACA GCT CCA TCA 13 TAC AGT CCC ACT GCG CCC TCC 14 TAT AGT CCA ACA GCC CCT AGT 15 TAC TCT CCA ACG GCT CCA TCA 16 TAC TCT CCT ACG GCG CCG TCC 17 TAC TCC CCA ACT GCT CCT GCA 18 TAT TCA CCA ACA GCA CCA TCA 19 TAC AGC CCA ACT GCA CCA TCT 20 TAT TCA CCA ACC GCT CCA TCT 21 TAT TCT CCA ACC GCC CCA AGC 22 TAC TCC CCA ACA GCC CCA AAC 23 TAC TCT CCT ACA GCG CCT AGC 24 TAC TCC CCA ACT GCC CCA GGT 25 TAC TCT CCT GGT GCC CCT GCT 26 TAT TCA CCA AAG CAA GAC GAA 27 CAA AAA CAC AAC GAG AAT GAA AAC TCC AGA	1 FSPTAPT 2 YSPTAPA 3 YSPTAPS 4 YSPTAPS 5 YSPTAPS 6 YSPTAPS 7 YSPTAPS 8 YSPTAPS 9 YSPTAPS 10 YSPTAPS 11 YSPTAPS 12 YSPTAPS 13 YSPTAPS 14 YSPTAPS 15 YSPTAPS 16 YSPTAPS 17 YSPTAPA 18 YSPTAPS 19 YSPTAPS 20 YSPTAPS 21 YSPTAPS 22 YSPTAPN 23 YSPTAPS 24 YSPTAPG 25 YSPGAPA 26 YSPKQDE 27 QKHENENSR
CTDS7A	1 TTT TCT CCA ACA TCT CCC ACC 2 TAC TCT CCT ACA AGT CCT GCT 3 TAC TCT CCA ACA AGC CCT GCT 4 TAT TCA CCC ACA TCC CCC GCC 5 TAT TCT CCT ACA TCT CCT GCA 6 TAC TCC CCA ACT TCT CCA GCC 7 TAT TCA CCG ACG TCA CCA GCC 8 TAC TCG CCT ACG AGC CCG GCG 9 TAC TCC CCA ACT TCC CCA GCA 10 TAT TCT CCA ACT AGT CCA GCC 11 TAC TCA CCT ACC AGT CCA GCA 12 TAT TCG CCA ACT TCT CCA GCC 13 TAT TCA CCA ACC TCT CCA GCA 14 TAC TCC CCA ACA TCT CCA GCG 15 TAC TCC CCA ACC TCG CCA GCA 16 TAC TCT CCG ACC AGT CCA GCC 17 TAC TCG CCT ACT TCC CCA GCT 18 TAC TCA CCA ACC TCT CCC GCA 19 TAT TCC CCT ACT TCA CCT GCC 20 TAT TCT CCT ACT TCC CCT GCA 21 TAT TCT CCT ACT TCT CCA GCC 22 TAC TCT CCC ACA TCG CCG AAC 23 TAT AGT CCT ACA TCG CCA GCT 24 TAT TCT CCC ACC TCT CCA GGA 25 TAC AGC CCG GGG TCT CCT GCT 26 TAT TCG CCT AAG CAA GAT GAA 27 CAA AAG CAT AAT GAA AAT TCC AGA	1 FSPTSPT 2 YSPTSPA 3 YSPTSPA 4 YSPTSPA 5 YSPTSPA 6 YSPTSPA 7 YSPTSPA 8 YSPTSPA 9 YSPTSPA 10 YSPTSPA 11 YSPTSPA 12 YSPTSPA 13 YSPTSPA 14 YSPTSPA 15 YSPTSPA 16 YSPTSPA 17 YSPTSPA 18 YSPTSPA 19 YSPTSPA 20 YSPTSPA 21 YSPTSPA 22 YSPTSPN 23 YSPTSPA 24 YSPTSPG 25 YSPGSPA 26 YSPKQDE 27 QKHENENSR

**Table S2.7 Oligonucleotides used in this study.**

Related to STAR Methods.

Name	Sequence	Source
Oligos used to clone gDNA targetting SEN1 in pFR622 (used to introduce the R308W mutation)	Forward: GATCATTCTGGTCGAGATTACTACGTTTTAGAGCTAG	This study
	Reverse: CTAGCTCTAAAACGTAGTAATCTCGACCAGAAT	This study
Oligos used to generate a donor DNA introducing the R308W mutation in SEN1	Forward: TCCGGCAAACCTGGACAGAGATTGTGGTATCTCAATTCTGGTCGTGGTTACTAC CCGTTTTCAATTTATTTGATAAAGACGTCTTCATAGA	This study
	Reverse: TCTATGAAGACGTCTTTATCAAATAAATTGAAAACCTGGTAGTAACCACGACCA GAATTGAGATAACCACAATCTCTGTCCAGTTTGCCGGA	This study
<i>SNR13</i> Northern blot probe	Forward: TTATAAATGGCATCTCAAATCGTC	Kim et al., 2006
	Reverse: CTGTCGCTCCGTGTCTCTTGTCCTG	Kim et al., 2006
<i>SNR47</i> Northern blot probe	Forward: CCTATAACAACAACAACATG	Tudek et al., 2014
	Reverse: ATAGCCATTAGTAAGTACGC	Tudek et al., 2014
NET-seq primers		
DNA linker	/5rApp/CTGTAGGCACCATCAAT/3ddC/	Ingolia et al., 2012
RT primer	/5Phos/AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCCG C/iSp18/CATCTA/iSp18/TTCAGACGTGTGCTCTTCCGATCTATTGATGGTGCCTA CAG	Ingolia et al., 2012
Library Forward PCR primer	AATGATACGGCGACCACCGAGATCTACAC	Ingolia et al., 2012
Library Reverse PCR primer (N701)	CAAGCAGAAGACGGCATAACGAGATTAAGGCGAGTACTGGAGTTCAGACGTG TGCTCTTCCG	Ingolia et al., 2012
Library Reverse PCR primer (N702)	CAAGCAGAAGACGGCATAACGAGATCGTACTAGGTACTGGAGTTCAGACGTG TGCTCTTCCG	Ingolia et al., 2012
Library Reverse PCR primer (N703)	CAAGCAGAAGACGGCATAACGAGATAGGCAGAAAGTACTGGAGTTCAGACGTG TGCTCTTCCG	Ingolia et al., 2012
Library Reverse PCR primer (N704)	CAAGCAGAAGACGGCATAACGAGATTCCTGAGCGTACTGGAGTTCAGACGTG TGCTCTTCCG	Ingolia et al., 2012
Read 1 primer HP10	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	Illumina
Index read primer	GATCGGAAGAGCACACGTCTGAACTCCAGTCAC	Illumina

**CHAPITRE 3: La stabilité de l'hélicase  
Sen1 est régulée par la phosphatase Glc7  
de façon dépendante du protéasome**

### 3.1 Résumé

#### **La stabilité de l'hélicase Sen1 est régulée par la phosphatase Glc7 de façon dépendante du protéasome**

Pierre Collin et François Robert

La terminaison transcriptionnelle chez *Saccharomyces cerevisiae* s'effectue à l'aide de deux mécanismes impliquant des complexes protéiques différents dépendamment des classes de gène. Aux gènes codants, le complexe de terminaison CPF-CF est recruté à la machinerie transcriptionnelle à l'aide d'interaction avec l'ARN naissant et avec la Ser2 phosphorylée du domaine en C-terminal (CTD) de l'ARN polymérase II. Aux gènes non-codants, comme les snoRNA et les CUT, l'ARN naissant et le CTD recrutent le complexe Nrd1-Nab3-Sen1 (NNS). Plusieurs évidences suggèrent que des composantes du CPF-CF, plus spécifiquement la phosphatase Glc7, sont aussi impliquées dans la terminaison par le NNS. Glc7 peut phosphoryler Sen1 *in vitro*, mais jusqu'à maintenant, le rôle biologique de cette activité n'est pas encore connu. Dans cette étude, nous démontrons que Glc7 régule la stabilité de Sen1. Lors de l'épuisement nucléaire de Glc7, le niveau de Sen1 diminue rapidement de façon dépendante du protéasome. Ces résultats fournissent, donc, un mécanisme pour le rôle précédemment démontré de Glc7 dans la terminaison des gènes non-codants. Fait intéressant, cette fonction de Glc7 est indépendante de son interaction avec le CPF-CF ou l'APT.

### 3.2 Contribution des auteurs

Conceptualisation: Pierre Collin et François Robert

Réalisation des expériences: Pierre Collin

Écriture du manuscrit: Pierre Collin et François Robert

Supervision: François Robert

### 3.3 Manuscript

#### Sen1 stability is regulated by Glc7 via a proteasome-dependent pathway

Pierre Collin<sup>1</sup> and François Robert<sup>1,2,3\*</sup>

<sup>1</sup> Institut de recherches cliniques de Montréal, 110 Avenue des Pins Ouest, Montréal, Québec, Canada, H2W 1R7.

<sup>2</sup> Département de médecine, Faculté de médecine, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, Canada, H3T 1J4.

<sup>3</sup> Lead contact

Contact information:

\*Correspondence: [francois.robert@ircm.qc.ca](mailto:francois.robert@ircm.qc.ca)

#### 3.3.1 Abstract

Transcriptional termination in *Saccharomyces cerevisiae* proceeds via two different mechanisms -each involving different protein complexes- depending on the gene class. At protein-coding genes, the CPF-CF termination complex is recruited to the transcription machinery via interactions with the nascent RNA and with the phosphorylated Ser2 of the RNA polymerase II C-terminal domain (CTD). For termination of non-coding RNAs such as snoRNAs and CUTs, the nascent RNA and the CTD recruit the Nrd1-Nab3-Sen1 (NNS) complex. Mounting evidence suggests that components of CPF-CF, more specifically the Glc7 phosphatase, are also involved in NNS termination. Glc7 can dephosphorylate Sen1 *in vitro*, but to date, the biological role of this activity has not been established. Here, we demonstrate that Glc7 regulates the stability of Sen1 via a proteasome-dependent mechanism. Upon nuclear depletion of Glc7, the level of Sen1 quickly declines in a proteasome-dependent manner. This provides a mechanism for the previously shown role of Glc7 in non-coding RNA termination. Interestingly, however, this function of Glc7 is independent of its interaction with CPF-CF or APT.

**Keywords:** Transcription termination, Glc7, Sen1, Ref2, Syc1, APT, CPF-CF, NNS

### 3.3.2 Introduction

Transcription can be divided into three stages: initiation, elongation and termination. Despite being very well studied, the last step remains the least understood. In *Saccharomyces cerevisiae*, two different mechanisms mediate termination at protein-coding genes and non-coding RNAs (Kuehner et al., 2011; Porrua and Libri, 2015; Richard and Manley, 2009). At protein-coding genes, termination is tightly coupled to the 3' end processing of the transcripts. The mechanism involves the recruitment of the CPF-CF complex at the polyadenylation (poly(A)) site followed by cleavage of the pre-mRNA by the endonuclease Ysh1 and the addition of a poly(A) tail by the poly(A) polymerase Pap1 (Zhelkovsky et al., 2006). In addition, the recruitment of CPF-CF induces the release of elongation factors, eventually leading to RNA polymerase II (RNAPII) termination (allosteric model) (Logan et al., 1987). In addition, the newly formed 5' end of the cleaved nascent RNA is targeted by the exonuclease Rat1 which is thought to chew its way up to the polymerase and induce termination (torpedo model) (Kim et al., 2004b; Sadowski et al., 2003). Initially thought to be exclusive mechanisms, evidence tends to show that these two mechanisms are complementary and cooperates to promptly induce termination (Dengl and Cramer, 2009; Luo et al., 2006). At short non-coding RNAs such as small nucleolar RNAs (snoRNAs) and cryptic unstable transcripts (CUTs), but also at some protein-coding genes (Arigo et al., 2006; Schulz et al., 2013; Thiebaut et al., 2006), termination occurs via a different mechanism that relies on Nrd1, Nab3 and Sen1 (also known as the NNS complex). Once recruited to the 5' end of the genes, it is proposed that the Sen1 helicase targets the R-loop formed by the nascent RNA and the template DNA to wrench its way up to RNAPII and to induce termination (Han et al., 2017; Hazelbaker et al., 2013; Porrua and Libri, 2013).

RNAPII contains a C-terminal domain (CTD) that is proposed to be the platform needed to recruit some of these termination factors (Kuehner et al., 2011). This domain is highly conserved and essential, and consists of tandem repetitions of the heptapeptide  $Y_1S_2P_3T_4S_5S_7$ . The size of the CTD ranges from 26 repetitions in *S. cerevisiae* to 52 in mammals (Eick and Geyer, 2013; Jeronimo et al., 2013). The modulation of the phosphorylation of Tyr, Ser and Thr CTD residues is involved in the recruitment of various



factors during transcription. During termination of protein-coding genes, the phosphorylation of Ser2 and Thr4 is involved in the recruitment of Pcf11 (CPF-CF complex) and Rtt103 via their C-terminal interaction domain (CID) (Collin et al., 2018; Harlen et al., 2016; Jasnovidova et al., 2017; Kim et al., 2004a; Kim et al., 2004b; Licatalosi et al., 2002; Lunde et al., 2010; Nemeč et al., 2017). Structural data from these CIDs and CTD also demonstrated that Tyr1 phosphorylation impairs interaction between these factors and the CTD (Meinhart and Cramer, 2004). Thus, it has been proposed that during elongation, the phosphorylated Tyr1 prevents early recruitment of CPF-CF complex (Mayer et al., 2012; Schreieck et al., 2014). Once Tyr1 is dephosphorylated at the poly(A) site, its hydroxyl group forms a hydrogen bond with the CID, which stabilizes the interaction. However, this model is controversial (Collin et al., 2018). Indeed, it has recently been demonstrated that mutation of Tyr1 or depletion of its phosphatase Glc7 only affects the termination of a limited number of genes (approximately 9%).

During NNS-dependent termination, phosphorylation of the CTD Ser5 plays a role in the recruitment of Nrd1 (Porrúa and Libri, 2015). This model is mainly based on *in vitro* evidences, as the structure of Nrd1 CID with the CTD, and on the correlation between Nrd1 and CTD Ser5-P occupancies (Kubicek et al., 2012; Mayer et al., 2012; Tudek et al., 2014; Vasiljeva et al., 2008). However, the role of CTD in Nrd1 recruitment is also controversial, as different studies have shown that the use of CTD S5A mutants and *nrd1*ΔCID mutant do not cause major termination defects (Collin et al., 2018; Vasiljeva et al., 2008). In addition, Nrd1 recruitment was shown to be unaffected in the *nrd1*ΔCID mutant. These results may be explained by the presence of redundant interactions such as the interactions between Nrd1 and the flat loop of Rpb2 and with the nascent RNA (Carroll et al., 2007; Creamer et al., 2011; Pearson and Moore, 2014). Surprisingly, Tyr1 appears to play a major role in NNS termination via a mechanism that act downstream of the recruitment of NNS factors (Collin et al., 2018). It has been proposed that Tyr1 is involved in pausing of RNAPII in the 5' region and that this is leveraged by Sen1, a poor processive helicase (Han et al., 2017; Porrúa and Libri, 2013), to catch RNAPII (Collin et al., 2018).

There is also some evidence for a role of some components of the CPF-CF complex in NNS termination (Grzechnik et al., 2015; Kim et al., 2006; Nedea et al., 2003; Steinmetz and Brow, 2003). Indeed, both phosphatases from CPF-CF, Glc7 and Ssu72, are involved in this process, but the mechanism is not well understood. It has been proposed that Glc7 is involved in the dephosphorylation of Sen1 to allow its activation (Nedea et al., 2008). Glc7 (and Ssu72) is part of the phosphatase module of CPF-CF (Nedea et al., 2008), which was recently shown to exist as a separate entity called APT (associated with Pta1) (Lidschreiber et al., 2018). APT, which is made of 7 subunits, is identical to the phosphatase module of CPF-CF with the exception of Syc1, which is present in APT but not included in CPF-CF complex. Although APT was shown to play a role in snoRNAs transcription, it does not appear to play a role in their termination (Lidschreiber et al., 2018). Hence, it remains unclear in which biochemical context Glc7 exerts its function in NNS termination.

In addition to be a constituent of CPF-CF and APT, Glc7 can also interact with NNS factors (Nedea et al., 2008). In addition, Glc7 can dephosphorylate Sen1 *in vitro* (Nedea et al., 2008). For these reasons, we investigated the role of Glc7 in the fate of Sen1 *in vivo*. Our results demonstrate that the dephosphorylation of Sen1 by Glc7 is necessary for its stability. In the absence of Glc7, Sen1 is degraded via a proteasome-dependent pathway and termination defects are observed at snoRNAs. Interestingly, this role of Glc7 is independent of its interaction with the CPF-CF complex or the APT, suggesting that it may function as part of a third entity.

### **3.3.3 Results**

#### **3.3.3.1 Glc7 regulates the stability of Sen1**

It has been shown that Sen1 is a substrate of Glc7 *in vitro* (Nedea et al., 2008). However, the potential function of this activity has never been investigated *in vivo* although it has been proposed that this activity may regulate the helicase activity of Sen1 (Nedea et al., 2008). In order to investigate the role of Sen1 dephosphorylation by Glc7, we performed Glc7 anchor-away followed by an immunoblotting of Sen1. Because Sen1 protein level is very low,

we tagged the endogenous *SEN1* gene with a triple Myc epitope and used immunoprecipitation to enrich Sen1 before detection by western blotting. This allowed us to avoid overexpression conditions and to perform all our experiments in physiological conditions. In wild type (WT) conditions, Sen1 is a very stable protein as shown by cycloheximide time course (**Figures 3.1A and 3.1B**). This stability, however, is greatly compromised when Glc7 is depleted from the nucleus using the anchor-away system (Haruki et al., 2008) (**Figures 3.1A and 3.1B**). Furthermore, this decrease of Sen1 is not observed the RNA level (**Figure 3.1C**) suggesting that Glc7 regulates Sen1 protein level.

To determine whether the effect of Glc7 on Sen1 protein requires the phosphatase activity of Glc7, we used two previously described catalytic mutants. E101Q was shown to have partial activity while D94N is catalytically dead (Logan et al., 2008). These mutants were used in a system of complementation after nuclear depletion of Glc7. The mutants were integrated at the *LEU2* locus in a strain where the endogenous *GLC7* gene is tagged FRB, so that Glc7 can be anchored-away from the nucleus upon addition of rapamycin. The mutants show growth defects that when plated on rapamycin-containing media, with D94N showing the strongest defect, as expected (**Figure 3.1E**). Interestingly, D94N has a growth defect even in the absence of rapamycin, suggesting some dominant-negative effect. When looking at Sen1 protein levels, we noticed decreased abundance in both mutants, with D94N showing the most dramatic effect. Hence, there is a correlation between Glc7 catalytic activity, the effect of the mutants on cell growth and the effect is has on Sen1 stability (**Figure 3.1E**). We therefore conclude that the stability of Sen1 is regulated by the phosphatase activity of Glc7. Interestingly, the effect of Glc7 nuclear-depletion on Sen1 can be rescued a co-treatment with proteasome inhibitors (**Figure 3.1F**). Finally, the depletion of Ssu72 had no effect on Sen1 protein levels, demonstrating that the effect of Glc7 on Sen1 is specific (**Figure 3.1G**). Taken together, these results demonstrate that the phosphatase activity of Glc7 regulates the stability of Sen1 via a proteasome-dependent manner.

### 3.3.3.2 Glc7 regulates Sen1's fate independently of its interaction with CPF-CF or APT

Glc7 is part of two separate complexes, CPF-CF and APT (Lidschreiber et al., 2018). In order to determine whether the action of Glc7 requires an interaction with one of these two complexes, we performed the auxin-induced degradation of Ref2 and Syc1 (**Figure 3.2**). Ref2 is required for the association of Glc7 with both complexes (Nedea et al., 2008). It is well established that in *ref2Δ* cells, Swd2 and Glc7 are no longer associated with CPF-CF (Nedea et al., 2008). It is important to note that the sole role of Swd2 (and Ref2) is thought to be the anchoring of Glc7 to CPF-CF and APT (Nedea et al., 2003; Nedea et al., 2008). Syc1, on the other hand, is required for the integrity of APT and plays no role with CPF-CF (Lidschreiber et al., 2018). Thus, degradation of Ref2 should lead to loss of Glc7's functions as part of both CPF-CF and APT, while degradation of Syc1 should only affect the function of Glc7 as part of APT. Interestingly, we found that deletion or neither had any substantial impact on Sen1 protein levels. Indeed, while both Ref2 and Syc1 are quickly decrease by ~90% 15 min after of auxin treatment, these depletions have no impact on Sen1 protein levels (**Figure 3.2**). These results strongly suggest that the role of Glc7 in Sen1's fate is independent of CPF-CF and APT complexes.

### 3.3.4 Discussion

Recently, it has been shown that the level of Sen1 is regulated during cell cycle via a proteasome-dependent pathway (Mischo et al., 2018). However, this study was performed under conditions of Sen1 overexpression to overcome technical difficulties related to its low protein level. Here, we confirm that Sen1 is degraded under physiological condition via a proteasome-dependent pathway. Importantly, we identified Glc7 as the phosphatase involved in regulating Sen1 stability. Interestingly, a proteome-wide analysis of phosphorylation and ubiquitination has identified a potential site in the N-terminal region of Sen1 (serine 863) that may work as a phosphodegron (Swaney et al., 2013). Interestingly, this region is conserved in Senataxin (SETX), the homologue of Sen1 in humans. It would be interesting to test whether this residue functions as a phosphodegron in yeast, and, in the event that it is, to validate if it plays the same role in humans. As SETX is involved in serious diseases such as amyotrophic

lateral sclerosis type 4, it would be very relevant to perform these experiments and to look if this potential site is mutated in disease samples (Blair et al., 2000).

Another important aspect of this study relates to the role of Glc7 in non-coding RNA termination. In addition to Sen1, Glc7 dephosphorylates the tyrosines within the RNAPII CTD (CTD Tyr1) (Schreieck et al., 2014). Recently, we showed that CTD Tyr1 is involved in termination of non-coding genes via NNS (Collin et al., 2018). In this work we took advantage of a strain where the Tyrosines in the CTD are mutated to non phosphorylatable (Phe) or phosphomimetic (Asp or Glu) residues. All three mutants showed strong termination defects at snoRNAs and CUTs, demonstrating the importance of Tyr1 -and suggesting a role of the tyrosine phosphorylation/dephosphorylation- in this process. A role for the phosphorylation/dephosphorylation was further supported by the fact that similar termination defects were observed upon anchor-away of Glc7. Because Glc7 has other substrates, however, prevented us from undoubtedly conclude about a role of dephosphorylation of Tyr1 by Glc7 in NNS termination. The present study, however, provides additional support for this model. Indeed, while the degradation of Ref2 is known to cause termination defects at snoRNAs, we showed that it does not cause the degradation of Sen1. Since the role of Ref2 is to allow the association of Glc7 with the CPF-CF or APT complex, it appears highly likely that the lack of termination caused by a deletion of *REF2* is caused by the inhibition of Tyr1 dephosphorylation (Nedea et al., 2008). To confirm this, the occupancy of Tyr1 should be determined in the strain Ref2-AID treated with auxin. If it turns out that the phosphorylation of Tyr1 is stabilized and a ncRNA termination defect is observed, the results would confirm the initial hypothesis that the dephosphorylation of Tyr1 is required for termination at non-coding genes. This could also be consolidated by the use of a Sen1 phosphodegron mutant preventing its degradation by Glc7 anchor-away.

Finally, once the different hypotheses validated, it will be possible to imagine a model involving Glc7 in three different complexes regulating the expression of non-coding genes (Lidschreiber et al., 2018; Nedea et al., 2008). Indeed, when in complex with CPF-CF, it would be involved in termination by the dephosphorylation of Tyr1. When in complex with

APT, it would be involved in transcriptional regulation of snoRNAs and when in complex with the NNS it would regulate the stability of Sen1.

### 3.3.5 Acknowledgments

This work was funded by grants from the Canadian Institutes of Health Research (CIHR) to F.R. (MOP-133648). F.R. holds a Research Chair from the « Fonds de Recherche Québec –Santé » (FRQS). P.C. held a studentship from the FRQS. We also thank P. Cramer for sharing yeast strains.

### 3.3.6 Reference

- 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.
- Blair, I.P., Bennett, C.L., Abel, A., Rabin, B.A., Griffin, J.W., Fischbeck, K.H., Cornblath, D.R., and Chance, P.F. (2000). A gene for autosomal dominant juvenile amyotrophic lateral sclerosis (ALS4) localizes to a 500-kb interval on chromosome 9q34. *Neurogenetics* 3, 1-6.
- Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* 13, 361-373.
- Collin, P., Jeronimo, C., Poitras, C., and Robert, F. (2018). RNA Polymerase II CTD Tyrosine 1 is Required for Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. *Mol Cell* *in press*.
- Creamer, T.J., Darby, M.M., Jamonnak, N., Schaugency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet* 7, e1002329.
- Dengl, S., and Cramer, P. (2009). Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* 284, 21270-21279.
- Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. *Chem Rev* 113, 8456-8490.

Grzechnik, P., Gdula, M.R., and Proudfoot, N.J. (2015). Pcf11 orchestrates transcription termination pathways in yeast. *Genes Dev* 29, 849-861.

Han, Z., Libri, D., and Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res* 45, 1355-1370.

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., and Churchman, L.S. (2016). Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* 15, 2147-2158.

Haruki, H., Nishikawa, J., and Laemmli, U.K. (2008). The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes. *Mol Cell* 31, 925-932.

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

Jasnovidova, O., Krejcikova, M., Kubicek, K., and Stefl, R. (2017). Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO Rep* 18, 906-913.

Jeronimo, C., Bataille, A.R., and Robert, F. (2013). The writers, readers, and functions of the RNA polymerase II C-terminal domain code. *Chem Rev* 113, 8491-8522.

Jeronimo, C., Watanabe, S., Kaplan, C.D., Peterson, C.L., and Robert, F. (2015). The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations. *Mol Cell* 58, 1113-1123.

Kim, M., Ahn, S.H., Krogan, N.J., Greenblatt, J.F., and Buratowski, S. (2004a). Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J* 23, 354-364.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedeá, E., Greenblatt, J.F., and Buratowski, S. (2004b). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517-522.

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.

Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev* 26, 1891-1896.

Kuehner, J.N., Pearson, E.L., and Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* 12, 283-294.

Laughery, M.F., Hunter, T., Brown, A., Hoopes, J., Ostbye, T., Shumaker, T., and Wyrick, J.J. (2015). New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. *Yeast* 32, 711-720.

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., and Bentley, D.L. (2002). Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Mol Cell* 9, 1101-1111.

Lidschreiber, M., Easter, A.D., Battaglia, S., Rodriguez-Molina, J.B., Casanal, A., Carminati, M., Baejen, C., Grzechnik, P., Maier, K.C., Cramer, P., *et al.* (2018). The APT complex is involved in non-coding RNA transcription and is distinct from CPF. *Nucleic Acids Res.*

Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr., and 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, 8306-8310.

Logan, M.R., Nguyen, T., Szapiel, N., Knockleby, J., Por, H., Zadworny, M., Neszt, M., Harrison, P., Bussey, H., Mandato, C.A., *et al.* (2008). Genetic interaction network of the *Saccharomyces cerevisiae* type 1 phosphatase Glc7. *BMC Genomics* 9, 336.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17, 1195-1201.

Luo, W., Johnson, A.W., and Bentley, D.L. (2006). 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.

Mayer, A., Heidemann, M., Lidschreiber, M., Schrieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336, 1723-1725.

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



Mischo, H.E., Chun, Y., Harlen, K.M., Smalec, B.M., Dhir, S., Churchman, L.S., and Buratowski, S. (2018). Cell-Cycle Modulation of Transcription Termination Factor Sen1. *Mol Cell* 70, 312-326 e317.

Morawska, M., and Ulrich, H.D. (2013). An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* 30, 341-351.

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

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol Cell* 29, 577-587.

Nemec, C.M., Yang, F., Gilmore, J.M., Hintermair, C., Ho, Y.H., Tseng, S.C., Heidemann, M., Zhang, Y., Florens, L., Gasch, A.P., *et al.* (2017). Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner. *Proc Natl Acad Sci U S A* 114, E3944-E3953.

Pearson, E., and Moore, C. (2014). The evolutionarily conserved Pol II flap loop contributes to proper transcription termination on short yeast genes. *Cell Rep* 9, 821-828.

Porrúa, O., and 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., and 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.

Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev* 23, 1247-1269.

Sadowski, M., Dichtl, B., Hubner, W., and Keller, W. (2003). Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J* 22, 2167-2177.

Schneider, B.L., Seufert, W., Steiner, B., Yang, Q.H., and Futcher, A.B. (1995). Use of polymerase chain reaction epitope tagging for protein tagging in *Saccharomyces cerevisiae*. *Yeast* 11, 1265-1274.

Schrieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., and Passmore, L.A. (2014). RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7. *Nat Struct Mol Biol* 21, 175-179.

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

Steinmetz, E.J., and Brow, D.A. (2003). Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. *Mol Cell Biol* 23, 6339-6349.

Swaney, D.L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J., and Villen, J. (2013). Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat Methods* 10, 676-682.

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.

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

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and 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, 795-804.

Zhelkovsky, A., Tacahashi, Y., Nasser, T., He, X., Sterzer, U., Jensen, T.H., Domdey, H., and Moore, C. (2006). The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in *Saccharomyces cerevisiae*. *RNA* 12, 435-445.

### 3.3.7 Figures

Figure 3.1

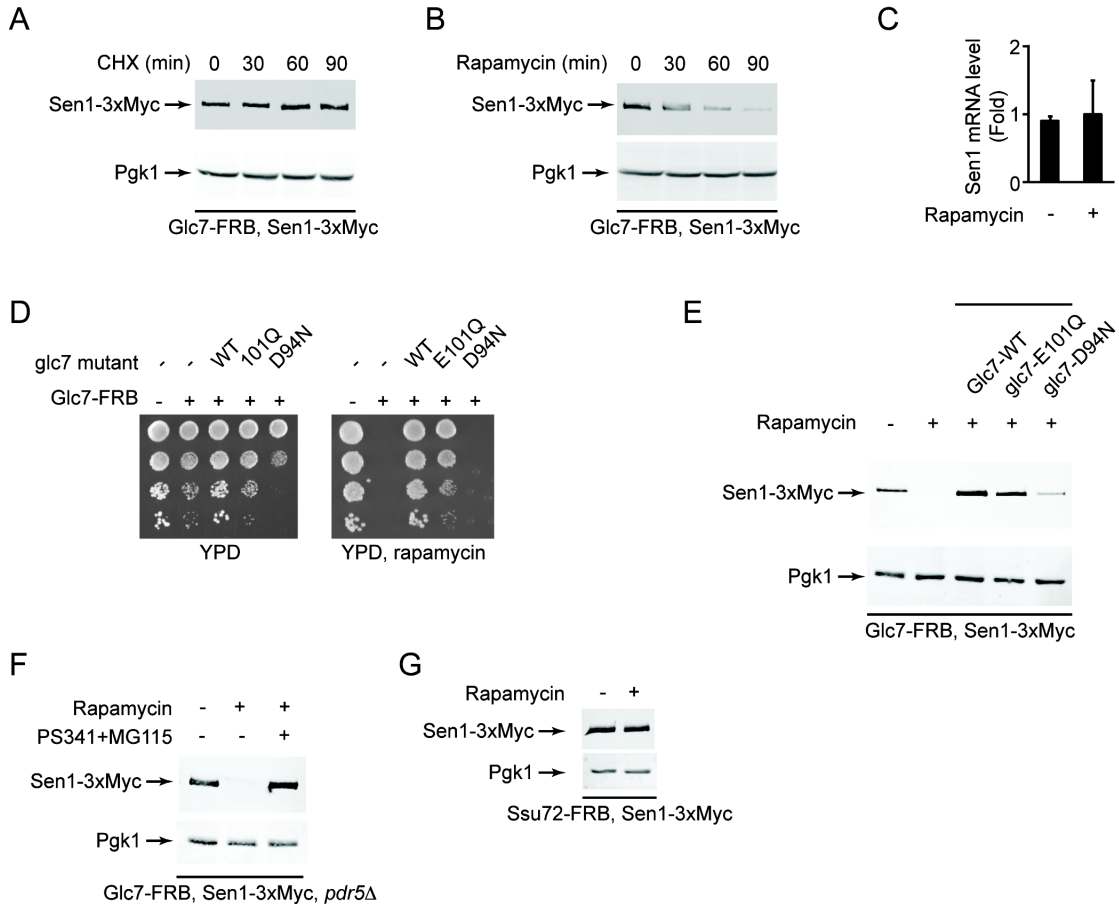
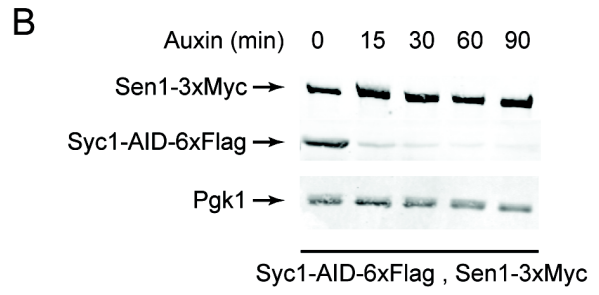
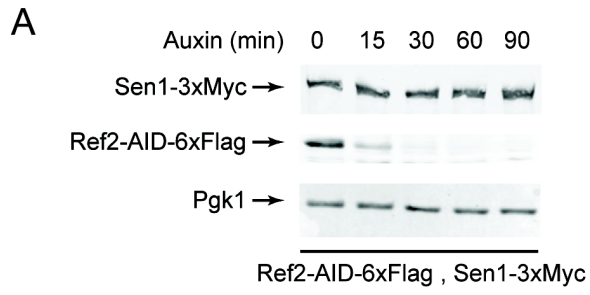


Figure 3.2



### 3.3.8 Figure legends

#### **Figure 3.1 Glc7-regulated Sen1 degradation via the proteasome.**

A) Immunoblot of Sen1 IPs upon a time course of translation inhibition using cycloheximide. Flowthrough extracts were used to immunoblot Pgk1. B) Immunoblot of Sen1 IPs upon a time course of Glc7 nuclear depletion by anchor-away. Flowthrough extracts were used to immunoblot Pgk1. C) Quantification of Sen1 mRNA level after nuclear depletion of the endogenous Glc7 protein by anchor-away (90 min) by RT-qPCR. D) Spotting assay for strains expressing the endogenous *GLC7*-WT-FRB and different *glc7*-mutant-3xHA integrated at the *LEU2* locus. Cells were spotted on YPD and YPD + rapamycin plates. E) Immunoblot of Sen1 IP after nuclear depletion of the endogenous Glc7 protein by anchor-away (90 min) and complemented with *GLC7*-WT-3xHA or *glc7*-mutant-3xHA integrated at the *LEU2* locus. Flowthrough extracts were used to immunoblot Pgk1. F) Immunoblot of Sen1 IP after depletion of the endogenous Glc7 protein by anchor-away (90 min) or co-treated with proteasome inhibitors. The *PDR5* gene was deleted to improve proteasome inhibition efficacy. Flowthrough extracts were used to immunoblot Pgk1. G) Immunoblot of Sen1 IP after depletion of the endogenous Ssu72 protein by anchor-away (90 min). Flowthrough extracts were used to immunoblot Pgk1.

#### **Figure 3.2 The role of Glc7 in Sen1 stability is independent of its interaction with CPF-CF or APT.**

A-B) Immunoblot of Sen1 IP upon a time course of auxin-induced degradation of Ref2 (A) or Syc1 (B). Flowthrough extracts were used to immunoblot Ref2, Syc1 and Pgk1.

### 3.3.9 Materials and Methods

#### Yeast culture and strains

Cells were inoculated at  $OD_{600}=0.1$  in 50 mL YPD supplemented with 2% glucose and grown at 30°C. Strains used in this study are listed in Table S3.1. Sen1, Ref2 and Syc1 tagging were performed by standard DNA recombination methods using the following plasmids: pMPY-3xMyc (Schneider et al., 1995) and pFA6a-AID-6xFLAG (Morawska and Ulrich, 2013). To construct *glc7* mutants, the *GLC7* gene (including its promoter and terminator) was cloned into pRS313 and 3xHA tag was added at the C-terminal using restriction enzyme-based cloning. Point mutations conferring reduced or catalytic dead activity were introduced by inverted PCR. Each construction was validated by immunoblotting and sequencing (plasmids are listed in Table S3.2). *GLC7* and *glc7* mutant genes (including promoter and terminator) from plasmids were amplified by PCR and recombined at the *LEU2* locus by CRISPR/Cas9 genome editing (Laughery et al., 2015). Briefly, a pair of oligonucleotides corresponding to a gRNA targeting a region into the ORF of *LEU2* (see Table S3.3) was designed for cloning into pML104 (URA3). Yeast cells were transformed with 400 ng of the gRNA-expressing plasmid and 1  $\mu$ g of a PCR product corresponding to the *GLC7* gene with flanking regions homologous to the *LEU2* locus, and plated on YNB –URA plates supplemented with 2% glucose. Clones were screened by PCR, sequencing and immunoblotting. The gRNA-expressing plasmid was then ejected from the selected clones by plating on YNB supplemented with 2% glucose and 1 mg/ml 5-Fluoroorotic acid.

#### Assays of Sen1 stability

When  $OD_{600}$  reached 0.4, cells were treated with 90  $\mu$ g/mL rapamycin for 90 min. Time course experiments were done with 90  $\mu$ g/mL rapamycin for 30, 60 and 90 min or with 500  $\mu$ M auxin for 15, 30, 60 and 90 min. For experiments involving the inhibition of the proteasome, the *PDR5* gene was deleted to improve inhibitor efficacy. Cells were treated with a cocktail of proteasome inhibitors (PS341 150  $\mu$ M and MG115 150  $\mu$ M) for 30 min before rapamycin treatment. After treatments, cells were washed with ice-cold Tris-buffered saline

(TBS; 50 mM Tris-HCl pH 7.5, 150 mM NaCl), resuspended in 700  $\mu$ L of Lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1 % Triton X-100, 0.1 % Na-deoxycholate) supplemented with protease inhibitors and lysed by bead beating. Lysates were centrifuged 5 min at 4°C and whole cell extracts were adjusted for each sample to have the same amount of OD units. For Sen1 IPs, 650  $\mu$ L of WCE were incubated with 50  $\mu$ L of Dynabeads coated with Pan Mouse IgG antibodies, pre-coupled with 2.5  $\mu$ L of anti-Myc antibody, overnight at 4°C with agitation. Beads were then washed six times with Lysis buffer. 50  $\mu$ L of 1X Laemmli buffer were added to beads, which were boiled 5 min and samples were separated on SDS-PAGE. Membranes were incubated with primary antibody (anti-Myc 9E10 (1:2000), anti-Pgk1 (1:2000) or anti-Flag M2 (1:2000)).

### **Serial-dilution growth assay**

Cells expressing the endogenous *GLC7*-FRB allele and any of the incorporated *GLC7*-3xHA (WT or mutant) alleles at the *LEU2* locus were grown to saturation, washed with H<sub>2</sub>O, diluted in water to OD<sub>600</sub>=8, followed by 10 fold serial dilutions. Cell dilutions were spotted on YPD or YPD + rapamycin (90  $\mu$ g/mL) plates, incubated at 30°C and inspected daily.

### **Sen1 mRNA expression level**

Cells were inoculated at OD<sub>600</sub>=0.1 in YPD supplemented with 2% glucose. When cells reached OD<sub>600</sub>=0.4, rapamycin was added at 90  $\mu$ g/mL and cells were further incubated for 90 min. Cells were washed with ice-cold water and RNA was extracted as in (Jeronimo et al., 2015) using the hot phenol extraction method. Total RNA was reverse-transcribed using Random primers and Superscripts III. Sen1 and control mRNA levels were quantified by qPCR and normalized with *ACT1* mRNA expression level.

### 3.3.10 Supplemented information

#### 3.3.10.1 Tables

**Table S3.1 Yeast strains used in this study.**

Strain	Alias	Genotype	Source
yFR116	Z1256	MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+	A kind gift of R.A. Young
yFR2564		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, Sen1-3MYC::URA3	This study
yFR3151		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, Sen1-3MYC::URA3, HIS3::promADH1-AFB2	This study
yFR3155		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, Sen1-3MYC::URA3, HIS3::promADH1-AFB2, REF2-AID-6FLAG::HphNT	This study
yFR3161		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL+, psi+, Sen1-3MYC::URA3, HIS3::promADH1-AFB2, SYC1-AID-6FLAG::HphNT	This study
yFR1321		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1	Euroscarf
yFR2347		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, Sen1-3MYC::URA3	This study
yFR1894		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6	Schrieck et al., 2014
yFR2348		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6, Sen1-3MYC::URA3	This study
yFR3096		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, Sen1-3MYC::URA3, pdr5Δ::LEU2	This study
yFR3096		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6, Sen1-3MYC::URA3, pdr5Δ::LEU2	This study
yFR3093		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6, LEU2::Glc7-3HA	This study
yFR3094		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6, LEU2::glc7-E101Q-3HA	This study
yFR3095		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6, LEU2::glc7-D94N-3HA	This study
yFR3100		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6, LEU2::Glc7-3HA, Sen1-3MYC::URA3	This study
yFR3101		MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1Δ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6, LEU2::glc7-E101Q-3HA, Sen1-3MYC::URA3	This study



yFR3102	MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, GLC7-FRB::KanMX6, LEU2::glc7-D94N-3HA, Sen1-3MYC::URA3	This study
yFR1895	MAT $\alpha$ , ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, GAL, psi+, tor1-1, fpr1 $\Delta$ ::NAT, RPL13A-2xFKBP12::TRP1, SSU72-FRB::KanMX6	Schrieck et al., 2014

**Table S3.2 Plasmids used in this study.**

<b>Plasmid</b>	<b>Description</b>	<b>Source</b>
pFA6a-FRB-kanMX6	FRB C-terminal tagging	Euroscarf
p3Flag-KanMX	3xFlag C-terminal tagging	Gelbart et al., 2001
pMPY-3xMyc	3xMyc C-terminal tagging	Schneider et al., 1995
pRS313	CEN, HIS3	Sikorski and Hieter, 1989
pFR666	CEN, HIS3, GLC7-3xHA	This study
pFR672	CEN, HIS3, glc7-E101Q-3xHA	This study
pFR673	CEN, HIS3, glc7-D94N-3xHA	This study
pRS303	HIS3	Sikorski and Hieter, 1989
pFR568	HIS3, pADH1-AFB2	Morawska and Ulrich, 2013

**Table S3.3 Oligonucleotides used in this study.**

Name	Sequence	Source
Oligos used to clone gDNA targetting <i>LEU2</i> in pFR622 (used to introduce the <i>glc7</i> mutants)	Forward: GATCCGTTGTTGTCAGAGAATTAG GTTT TAGAGCTAG	This study
	Reverse: CTAGCTCTAAAACCTAATTCTCTGA CAACAACG	This study
Oligos used to amplify <i>glc7</i> genes to generate a donor in <i>LEU2</i>	Forward: TTTACATTTTCAGCAATATATATATA TATATTTCAAGGATATAACCATTCTA GCTTTTGCGGTTGTTATTCA	This study
	Reverse: ATTTTCATTTATAAAGTTTATGTACA AATATCATAAAAAAAGAGAATCTT TCCAAAGGAAGACGTGACCAT	This study
RT-qPCR, <i>SEN1</i>	Forward: TGCTGTTCGTGTTTTGATGG	This study
	Reverse: TTGGTGAAATTTTCCAGTTCA	This study
RT-qPCR, <i>ACT1</i>	Forward: ACCATCACCGGAATCCAAAA	This study
	Reverse: TCGTGCTGTCTTCCCATCTAT	This study



# **CHAPITRE 4: Discussion et conclusion**

#### 4.1 Le rôle du CTD dans la terminaison des gènes codants

Les résultats obtenus lors de mon projet principal ont permis d'effectuer une mise à jour importante et de clarifier le rôle du CTD dans la terminaison transcriptionnelle, plus particulièrement celui du résidu Tyr1.

D'abord, nos résultats de ChIP-chip confirment l'implication de la Ser2 et de la Thr4 dans la terminaison des gènes codants. Principalement basés sur la détermination des structures de leurs domaines CID avec le CTD, les facteurs Rtt103 et Pcf11 ont été proposés pour être recrutés par la Ser2-P [98, 235]. Il a aussi été démontré récemment que la Thr4 joue un rôle important dans cette interaction avec Rtt103 [246, 247]. En effet, le groupement phosphate de la Thr4 est impliqué dans une interaction directe avec le CID lorsque la phosphorylation de la Ser2 est absente. Puis, lorsque la Ser2 est phosphorylée, le groupement hydroxyle de la Thr4 effectue un pont hydrogène avec le phosphate de la Ser2-P, ce qui stabilise la structure en coude  $\beta$  du CTD et, par le fait même, l'interaction avec le facteur. Des évidences génomiques ont aussi confirmé ce modèle, où l'utilisation d'un mutant T4V a démontré une diminution du recrutement de Rtt103 [134]. Pour ce qui est du recrutement de Pcf11, il est plutôt controversé. En effet, différentes études ont démontré que le recrutement de Pcf11 n'est pas aboli par deux mutants connus pour abroger son interaction avec le CTD [235, 521]. Ces résultats sur le recrutement de Pcf11 suggèrent donc que le CTD n'est pas le seul élément impliqué. En effet, il a aussi été démontré qu'il est recruté par la sous-unité Rpb2 de l'ARNPII [510]. De plus, comme Pcf11 est en interaction avec le CPF-CF, des motifs d'ARN sont aussi impliqués. Donc, le CTD n'est qu'une des trois interactions connues pour être impliquées dans le recrutement du facteur. Ainsi, une perte d'interaction avec le CTD n'abolit pas son recrutement.

Ces différentes observations ont été confirmées dans notre étude. En effet, une diminution de Rtt103 est observée en ChIP-chip avec les mutants S2A et T4V à l'aide du système CAND. Comme le traitement à la rapamycine n'est pas efficace à 100 %, il n'est pas possible d'obtenir une réduction complète du facteur ciblé. Pour ce faire, nous avons utilisé le mutant *ctk1Δ* et un mutant endogène S2A du CTD [611]. Dans les deux cas, le recrutement de

Rtt103 est aboli complètement, confirmant que la Ser2 est nécessaire à son recrutement. Il n'a cependant pas été possible d'effectuer les mêmes expériences avec la Thr4, car sa kinase n'est pas encore connue chez la levure. Cependant, comme le mutant T4V n'est pas létal, il serait envisageable d'effectuer une mutation endogène afin d'effectuer cette étude. Pour ce qui est de Pcf11, on observe une diminution de son recrutement avec les mutants S2A et T4V. Contrairement à Rtt103, une diminution partielle (et non complète) est observée avec le mutant *ctk1Δ* et le mutant endogène S2A du CTD. Ces résultats confirment que la Ser2 et fort probablement la Thr4 (basé sur les évidences structurales) sont impliquées dans le recrutement de Pcf11, mais que d'autres éléments partiellement redondants sont aussi impliqués, comme Rpb2 ou l'ARN naissant. Le rôle *in vivo* de la Thr4 dans le recrutement de Rtt103 a récemment été caractérisé, mais pour ce qui est de son implication dans le recrutement de Pcf11, il s'agit des premières évidences [134]. Enfin, ce recrutement partiel de Pcf11 avec les mutants S2A et T4V se reflète par une diminution de l'efficacité de terminaison qui provoque un défaut mineur s'étalant sur environ 100 bp, contrairement à la perte totale de Pcf11 qui peut entraîner des défauts de terminaison de plusieurs centaines de paires de base [516]. Il est d'ailleurs intéressant de noter que contrairement au mutant S2A, le mutant T4V n'est pas létal, et ce, même si la terminaison est globalement affectée. Cela suggère que la Ser2 est impliquée dans d'autres processus essentiels, indépendants de la Thr4.

Les résultats discutés plus haut confirment ce qui a été démontré dans la littérature. Toutefois, le rôle de la Tyr1 dans la terminaison des gènes codants, tel que suggéré par nos résultats, s'avère très différent de celui décrit dans la littérature. En effet, il est proposé que la déphosphorylation de la Tyr1 est nécessaire afin de permettre une interaction entre le CTD phosphorylé en Ser2 et Pcf11/Rtt103 [128, 142]. En présence de la Tyr1-P, il y aurait un encombrement stérique et l'interaction n'aurait pas lieu, évitant ainsi un recrutement précoce du complexe CPF-CF lors de l'élongation. Donc, si le modèle dit vrai, nous aurions dû observer une terminaison précoce dans le mutant Y1F. Cependant, ce n'est pas ce qui est observé en CHIP-chip et en NET-seq. En effet, on observe plutôt un défaut de terminaison (plutôt qu'une terminaison précoce), et ce à un groupe limité de gènes (environ 9 %). Afin de valider notre découverte, nous avons analysé de nouveau les résultats publiés sur la phosphatase de la Tyr1, Glc7, car le modèle est principalement basé sur cette étude [142].

Lorsque l'on regarde attentivement les différentes analyses des résultats de cette publication, on peut voir qu'elles ont toutes été effectuées à l'échelle du génome sous forme de *metagene* et que seulement quelques exemples de gènes spécifiques sont montrés. En effectuant ce genre d'analyse avec notre mutant Y1F, il est aussi possible d'observer un défaut de terminaison qui semble être général. Il faut donc être prudent avec les analyses de type *metagene*, car celles-ci peuvent parfois amplifier des effets sporadiques. Ainsi, en effectuant une analyse gène par gène avec les résultats publiés pour Glc7, on observe des défauts de terminaison gènes spécifiques, similaires à ceux observés avec le mutant Y1F. Des analyses approfondies ont ensuite été effectuées afin d'identifier quelles caractéristiques communes pouvaient avoir ces quelques gènes, mais rien n'a été trouvé. Enfin, ces résultats pour la Tyr1 aux gènes codants sont différents de ceux obtenus chez les mammifères [88]. En effet, il a été démontré chez l'humain que le résidu est impliqué globalement dans la terminaison des gènes codants. Cependant, lors de conversations privées avec le chercheur responsable de l'étude, Dirk Eick, et de l'auteur principale, Nilay Shah, ceux-ci ont affirmé que les résultats obtenus dans leur étude n'étaient pas si différents que ceux obtenus chez la levure, dans la mesure où c'est seulement 40 % des gènes codants qui dépendent de la Tyr1 et non la majorité.

#### **4.2 Le rôle du CTD dans la terminaison des gènes courts non-codants**

D'autres résultats étonnants ont aussi été obtenus pour la terminaison des ARN non-codants, impliquant cette fois-ci le complexe NNS. Selon le modèle accepté, la phosphorylation de la Ser5 dans la région 5' des gènes serait nécessaire afin de recruter le facteur Nrd1 [241]. Cependant, avec notre mutant S5A, aucun défaut de terminaison n'est observé par ChIP-chip ou par *Northern blot*. Il faut toutefois être prudent avec ce mutant, car des défauts d'échappement du promoteur sont présents, ce qui rend la détection de défaut de terminaison difficile [146]. Donc, afin de rectifier cette limitation, un mutant *nrd1ΔCID* a été utilisé en ChIP-chip. Si le modèle est vrai, on devrait aussi observer un défaut de terminaison avec ce mutant, mais ce n'est toutefois pas le cas. Étonnamment, une étude publiée en 2014 utilisant ce même mutant démontre un défaut de terminaison de l'ARNPII, mais après avoir observé attentivement ces résultats, on remarque que les défauts sont assez faibles [554]. De



plus, nos résultats confirment une étude publiée par le groupe de Stephen Buratowski, dans laquelle il n'observe pas de défaut de terminaison avec ce mutant en *Northern blot* [241]. Donc, bien qu'il soit probable que la Ser5-P joue un rôle dans la terminaison par le NNS, son rôle demeure marginal comparativement à celui de la Tyr1 (voir plus bas).

Des défauts de terminaison sont toutefois présents avec le mutant S2A aux snoRNA. Ces résultats confirment d'autres études qui ont observé le même phénomène avec le mutant *ctk1Δ* et le mutant endogène de la S2A du CTD [552]. Ces défauts peuvent être en partie expliqués par la diminution du recrutement de Pcf11, car il a aussi été proposé que ce facteur soit impliqué dans ce mécanisme, sans que l'on comprenne exactement son rôle [521, 550]. Ainsi, on devrait observer un défaut de terminaison partiel avec le mutant T4V, ce qui n'est pas le cas. La Ser2 est donc impliquée dans le recrutement d'autres facteurs, en plus de Pcf11, qui sont nécessaires à la terminaison de ces gènes. Il a été proposé que la Ser2-P est impliquée dans le recrutement de l'hélicase Sen1 par une interaction avec le résidu R302 [555, 556]. Cependant, un mutant de ce dernier (R302W) n'occasionne pas de défaut de terminaison ou même de défaut de recrutement du facteur en CHIP-chip. Donc, le recrutement de Sen1 semble être indépendant de son interaction avec le CTD. Il ne faut toutefois pas exclure la possibilité qu'un autre résidu soit impliqué, car un défaut de recrutement de Sen1 est visible avec un mutant Rpb1 sans CTD.

Le résultat le plus surprenant a été obtenu avec le mutant de la Tyr1. En effet, il est possible d'observer des défauts majeurs de la terminaison par CHIP-chip et par NET-seq à tous les gènes nécessitant le complexe NNS, soit les snoRNA, les CUT et les gènes codants régulés par atténuation. Ces résultats sont surprenants, car il n'avait jamais été proposé auparavant que la Tyr1 puisse être impliquée dans ce processus. L'une des premières questions qui se pose alors est : est-ce la phosphorylation, la déphosphorylation ou la tyrosine elle-même qui est nécessaire à la terminaison? Comme le mutant Y1F ne contient pas de groupement hydroxyle, il ne peut pas être utilisé pour répondre à la question. L'idéal aurait été d'utiliser un mutant de la kinase de la Tyr1, mais celle-ci n'est pas encore connue. Une autre possibilité est l'utilisation d'un mutant de sa phosphatase Glc7. Nos résultats démontrent que le *anchor-away* de Glc7 génère des défauts de terminaison qui mimique ceux observés avec le mutant

Y1F. Bien que ces résultats tentent à suggérer que la déphosphorylation de la Tyr1 soit requise pour la terminaison par le NSS, le fait que Glc7 possède d'autres substrats, incluant Sen1, limite l'interprétation de ces données [541]. Il serait donc utile de déterminer exactement le rôle de Glc7 dans la déphosphorylation de la Sen1 et de déterminer le résidu ciblé, ce qui est l'objectif de mon deuxième projet. Par contre, l'utilisation de mutants phosphomimétiques de la Tyr1 (Y1D et Y1E) supporte l'importance de la déphosphorylation de la Tyr1. Ces mutants provoquent des défauts d'initiation de la transcription, probablement causés par une perte d'interaction entre l'ARNPII et le Médiateur, comme il a récemment été proposé [88]. Cependant, malgré ce niveau d'initiation diminué, on a pu constater que ces mutants provoquent des défauts de terminaison similaires à ceux observés avec le mutant Y1F. Enfin, nous avons confirmé que la Tyr1 est bel et bien phosphorylé aux gènes snoRNA. Ces résultats sont aussi étonnants étant donné la faible taille de ces unités de transcription (moins de 100 bp) et que la phosphorylation de la Tyr1 est progressive durant l'élongation pour atteindre un plateau à environ 500bp. Il était donc assumé que la Tyr1 n'était pas phosphorylée à ces gènes. Donc, tous ces résultats suggèrent que la déphosphorylation de la Tyr1 est nécessaire afin d'induire la terminaison par le NNS. D'autres études devront toutefois être effectuées afin de consolider cette hypothèse.

Maintenant, quelle est la cause exacte du défaut de terminaison avec le mutant Y1F? L'hypothèse la plus simple, basée sur le modèle, est une perte de recrutement de Nrd1 et/ou de Sen1 dans le mutant Y1F. Toutefois, la distribution des deux facteurs, tel que déterminée par ChIP-chip, démontre que leur recrutement est indépendant de la Tyr1, voir même de leur interaction avec le CTD, comme le montre l'utilisation des mutants *nrd1*ΔCID et *sen1*-R302W. Il est toutefois important de noter que le mutant Rpb1 sans CTD provoque une diminution de leur recrutement, ce qui suggère que le domaine est tout de même impliqué indirectement. De plus, ces résultats de recrutement montrent un phénomène très intéressant. En effet, les facteurs recrutés semblent rester accrochés à l'ARNPII durant le défaut de terminaison. En effet, dans les souches sauvages, la distribution de Nrd1 et Sen1 forme un pic, mais lorsque les facteurs sont en présence du mutant Y1F, ils semblent trainer avec la polymérase. Ces résultats proposent que la Tyr1 du CTD est impliquée dans un événement se situant en aval du recrutement des deux facteurs. Récemment, il a été proposé que l'ARNPII

doit subir une pause ou un ralentissement pour permettre à Sen1, une hélicase très peu processive, de rattraper la polymérase et d'induire la terminaison [53, 547]. Ainsi, des expériences de NET-seq ont été effectuées afin de vérifier l'implication de pauses Tyr1-dépendentes.

Le NET-seq est une technique puissante pour différentes raisons. Elle est de haute résolution (un nucléotide) et les résultats obtenus sont brins spécifiques [432]. De plus, comme la résolution est très élevée, il est possible d'observer des événements de pauses qui ne sont pas visibles par ChIP-chip ou par ChIP-seq. Ainsi, l'analyse des résultats de NET-seq obtenus avec le mutant Y1F a permis de constater que des pauses dans la région proximale au promoteur sont diminuées de façon significative dans le mutant Y1F par rapport au sauvage. Même si les pauses de l'ARNPII sont beaucoup plus étudiées chez les métazoaires, celles-ci sont aussi présentes chez *S. cerevisiae*, plus précisément dans la région 5' des gènes [432]. Les résultats obtenus avec nos expériences confirment cette observation. De plus, en déterminant tous les sites de pauses présents chez la souche sauvage et en les comparant avec le mutant Y1F, il est possible de constater une diminution significative des pauses situées dans la région 5' des gènes par rapport à celles situées plus en aval dans les ORF. L'analyse des données de NET-seq aux snoRNA et aux CUT démontre qu'une pause dépendante de la Tyr1 est également présente à ces loci. Comme ces gènes sont très courts, la pause se situe à proximité du site de terminaison. Enfin, il est intéressant de noter que ces résultats pour la Tyr1 sont similaires à ceux observés chez les mammifères [88]. Ainsi, en nous basant sur ces observations, nous avons émis l'hypothèse que la Tyr1 est impliquée dans des événements de pause en 5' des gènes, et que ceux-ci sont exploités par l'hélicase Sen1 pour qu'elle puisse rattraper la polymérase et induire la terminaison.

Afin de tester ce modèle, deux stratégies connues pour induire des pauses ou des ralentissements de l'ARNPII ont été utilisées. La première consiste à traiter le mutant Y1F avec le 6-azauracile (6AU), connu pour induire une diminution globale du niveau cellulaire de GTP, provoquant des pauses ou un ralentissement de la polymérase [422, 423]. La deuxième stratégie consiste, quant à elle, à combiner la mutation Y1F avec des mutations de Rpb1 connues pour ralentir l'ARNPII, soit N488D ou H1085Y. Ces mutations affectent l'efficacité

d'incorporation du NTP au site catalytique de l'ARNPII, ce qui provoque un ralentissement de l'élongation [45, 53-55, 58, 59]. Les résultats obtenus en CHIP-chip avec ces deux stratégies sont très clairs : il est possible de corriger les défauts de terminaison aux snoRNA et aux CUT lorsque la polymérase Y1F subit un ralentissement, ce qui supporte notre hypothèse. De plus, il est aussi possible de corriger les défauts de terminaison aux gènes codants qui dépendent de la Tyr1 (9%), et ce, même s'ils sont indépendants du NNS pour leur terminaison et qu'aucun évènement de pause n'explique leur défaut de terminaison. Ce résultat demeure encore inexpliqué. Enfin, il est très intéressant de noter que ces stratégies corrigent également le défaut de croissance du mutant Y1F. Comme les défauts de terminaison de ce mutant ciblent principalement les snoRNA, il est possible qu'en corrigeant ces défauts, la synthèse et la maturation des ribonucléoprotéines soient rétablies. Il aurait donc été intéressant de vérifier si la maturation des ARN ribosomiaux est compromise avec le mutant Y1F.

Ces résultats sont par contre consistants avec une autre hypothèse : le mutant Y1F provoque une augmentation de la vitesse d'élongation, ce qui empêche Sen1 de rattraper la polymérase. Il a été démontré que l'utilisation d'un mutant rapide de l'ARNPII, *rpb1-E1103G*, provoque des défauts de terminaison. Il est cependant important de noter que cette proposition est basée sur une expérience de RT-qPCR à un seul snoRNA [53]. Donc, pour vérifier cette contre hypothèse, nous aurions pu mesurer la vitesse d'élongation du mutant Y1F [422]. Cette expérience consiste à utiliser un long gène (8 kb; YLR454W) sous le contrôle du promoteur *GALI*. Les cellules croient dans un milieu complété en galactose, ce qui induit le gène. Puis, lorsque la densité cellulaire voulue est atteinte, l'ajout de glucose est effectué, ce qui provoque la translocation au noyau du facteur Mig1 qui induit l'arrêt de la transcription du gène. Des expériences de CHIP sont alors effectuées à différents temps et à différentes positions du gène afin de déterminer la progression et la vitesse de la dernière vague des ARNPII. Cette expérience a été faite avec le mutant Y1F et les résultats obtenus sont très similaires à ceux obtenus avec le sauvage. Toutefois, une étude a récemment démontré une limite de ce système pour mesurer la vitesse [54]. En effet, ils ont démontré que le mutant *rpb1-E1103G*, connu pour être très rapide, apparaît lent lorsque l'on détermine sa vitesse avec ce système. Ce résultat peut être expliqué par une diminution de la translocation au noyau du facteur Mig1 lorsqu'il est en présence du mutant rapide, ce qui ne permet pas d'inhiber

totale de l'induction du gène. Ainsi, les résultats obtenus avec notre mutant ne peuvent pas être tenus en compte. Une autre possibilité pour effectuer une inhibition rapide de la transcription, indépendamment de Mig1, est un traitement à la triptolide [386]. Cependant, la drogue n'a aucun effet sur la levure, car elle est probablement chassée de la cellule à l'aide d'une pompe membranaire.

Enfin, pour tenter de vérifier cette contre hypothèse, nous avons opté pour une stratégie alternative. Comme l'hypothèse est basée sur l'utilisation d'un mutant rapide de la polymérase, nous avons utilisé le mutant *rpb1-G1097D* et nous avons déterminé sa distribution en ChIP-chip. Les résultats obtenus ne montrent aucun défaut de terminaison. De plus, il est important de noter que nous avons aussi confirmé ce résultat avec des données en ChIP-seq pour le mutant *rpb1-E1103G*, récemment publiées par Stephen Buratowski, du même groupe de recherche qui affirmait observer des défauts de terminaison avec ce mutant en RT-qPCR [53, 548]. Ces résultats permettent donc d'appuyer notre hypothèse et de conclure que la Tyr1 est impliquée dans des événements de pause nécessaires aux NNS afin d'induire la terminaison.

La prochaine étape sera donc de déterminer comment la Tyr1 peut induire des événements de pause dans la région 5'. Lors du projet, des expériences de protéomique avec le mutant Y1F ont été effectuées afin d'identifier des protéines dont l'interaction avec l'ARNPII est affectée dans ce mutant. De façon générale, rien de significatif n'est ressorti avec cette expérience du point de vue de la terminaison sauf peut-être un complexe, celui du coiffage. Comme il a été soulevé en introduction, le groupement hydroxyle de la Tyr1 est souvent nécessaire dans l'interaction du CTD avec les facteurs, ce qui est aussi le cas pour ce complexe. Il y a quelques années, des études ont démontré l'implication de la guanylyltransférase Ceg1 et de l'ARNm 5'-triphosphatase Cet1 dans des événements de pause dans des régions promotrices [612, 613]. Comme le recrutement du complexe est stabilisé par un pont hydrogène avec la Tyr1, il est possible que son recrutement soit affecté par le mutant Y1F, ce qui pourrait provoquer des diminutions de pauses. Ainsi, une expérience utilisant la technique *anchor-away* avec l'enzyme du coiffage suivie d'expériences de ChIP-chip de l'ARNPII pourraient être effectuées afin de vérifier si des défauts de terminaison sont

observés aux gènes non-codants. Si c'est le cas, des expériences de NET-seq pourront être effectuées afin de vérifier la perte de pauses. Enfin, le système développé durant cette étude pourrait aussi être utilisé pour étudier les autres étapes de la transcription (l'initiation et l'élongation) et des événements co-transcriptionnels, comme l'épissage, afin de valider la contribution de chacun des résidus du CTD.

### 4.3 Le rôle de la déphosphorylation de Sen1 par la phosphatase Glc7

Au fil des ans, l'utilisation de différents mutants des composantes du CPF-CF a permis de constater que le complexe (ou du moins certains de ses constituants) joue aussi un rôle important dans la terminaison des ARN non-codants [202, 521, 550, 551]. Le complexe contient quatre enzymes, soit les phosphatases Ssu72 et Glc7, l'endonucléase Ysh1 et la poly(A) polymérase Pap1. Il a été démontré que, de ces facteurs, seules les phosphatases sont impliquées dans la terminaison des gènes non-codants [521, 524, 541, 551]. Lors de mes études, j'ai eu l'occasion d'effectuer le *anchor-away* des facteurs Ssu72, Glc7 et Ysh1 afin d'observer la distribution de l'ARNPII en ChIP-chip. Des analyses aux ARN non-codants ont permis de valider l'implication globale de Ssu72 et de Glc7 à ces gènes, contrairement à Ysh1 (résultats non publiés). Dans la littérature, aucun mécanisme n'a été proposé pour Ssu72. Pour ce qui est de Glc7, il a été démontré qu'il peut effectuer la déphosphorylation de l'hélicase Sen1 *in vitro* [541]. En se basant sur cette observation, les auteurs ont ensuite proposé un modèle où Glc7 serait recruté au complexe NNS pour ensuite déphosphoryler Sen1 afin de permettre son activation. Toutefois, aucune évidence *in vivo* ne le démontre.

Dans le cadre de mon deuxième projet, les résultats obtenus permettent de proposer un mécanisme *in vivo* dans un contexte physiologique. En effet, comme Sen1 est très peu exprimé, il est souvent surexprimé ou étudié dans des contextes *in vitro*. Pour ce faire, nous avons effectué un *anchor-away* de Glc7 suivi d'une immunoprécipitation de Sen1. Les résultats obtenus ont ainsi permis de proposer un modèle où Glc7 serait responsable de la déphosphorylation de Sen1, non pas pour l'activer, mais pour le stabiliser. Dans le cas contraire, Sen1 est ciblé par le protéasome afin d'être dégradé. De plus, nos résultats montrent

que la dégradation inductible à l'auxine du facteur de terminaison Ref2 n'affecte pas la stabilité de Sen1. Comme le rôle de ce facteur est de permettre l'association de Glc7 aux complexes CPF-CF et APT, nos résultats permettent d'appuyer la proposition que Glc7 seul interagit avec le NNS [541]. Le même type d'expérience a aussi été effectué avec la protéine Syc1 du complexe APT afin de pouvoir discriminer le CPF-CF de l'APT, car son rôle est de maintenir l'intégrité de l'APT [513]. Comme attendu, aucun effet sur Sen1 n'a été observé.

Un détail important reste toutefois à déterminer : le site de phosphorylation qui correspond au phospho dégron de Sen1. Comme l'hélicase est très peu abondante (environ 300 molécules/cellule) et qu'il s'agit de la deuxième plus grosse protéine chez *S. cerevisiae*, l'identification de ce site par criblage consisterait en un défi technique et dispendieux. Il y a quelques années, une étude portant sur l'identification de sites de phosphorylation et d'ubiquitination à l'échelle du protéome chez la levure a permis d'identifier un site potentiel (S863) dans la région N-terminal de Sen1 [614]. Fait intéressant, la région où se situe ce site est conservée chez SETX, l'homologue de Sen1 chez l'humain. Comme SETX est impliquée dans différentes maladies graves, comme la sclérose latérale amyotrophique de type 4, il devient donc pertinent de valider ce site chez la levure et, ensuite, s'il s'agit du phospho dégron, il sera possible de vérifier dans différentes lignées cellulaires humaines si le même site permet de stabiliser l'hélicase [575]. Finalement, cette région pourrait être séquencée chez des patients atteints de ces maladies afin de déterminer s'il existe une corrélation avec la présence de mutations dans ce phospho dégron. Cependant, aucune mutation relative à cette sérine n'a été découverte jusqu'à maintenant. Enfin, il est intéressant de noter qu'une diminution significative de l'homologue de Glc7, PP1, provoque aussi des défauts de terminaison [615]. Il serait donc possible de vérifier l'impact d'une inhibition de PP1 sur SETX.

Lorsque l'identification du site de phosphorylation de Sen1 sera effectuée, il sera possible de muter ce résidu et de stabiliser Sen1. Cette expérience pourra ensuite être utilisée pour appuyer l'hypothèse de mon premier projet qui veut que la déphosphorylation de la Tyr1 du CTD par Glc7 soit nécessaire afin de permettre une terminaison efficace des ARN non-codants par le NNS. En effet, avec un tel mutant, il sera possible de cibler spécifiquement la phosphorylation de la Tyr1 lorsqu'un *anchor-away* de Glc7 sera effectué (l'effet sur Sen1

étant supprimé par la mutation de son phospho dégron). Ainsi, dans le cas où un défaut de terminaison serait toujours visible, notre hypothèse serait alors confirmée. Si, par contre, la mutation du phospho dégron de Sen1 devait prévenir les défauts de terminaison de l'*anchor-away* de Glc7, une implication de la stabilité de Sen1 par Glc7 dans la terminaison devrait alors être considérée. De plus, les résultats obtenus avec la dégradation de Ref2 inductible à l'auxine peuvent aussi être utilisés pour valider cette hypothèse. En effet, Ref2 a été caractérisé pour être impliqué dans la terminaison des gènes non-codants et dans la présente étude, pour ne pas être impliqué dans la stabilisation de Sen1 [202]. Comme le rôle de Ref2 se limite à permettre l'association de Glc7 au CPF-CF ou à l'APT, il s'agira de vérifier l'état de phosphorylation de la Tyr1 lorsque Ref2 est dégradé. Ainsi, si l'hypothèse est bonne, on devrait observer une phosphorylation constante du résidu et un défaut de terminaison. Un contrôle sera toutefois important afin de démontrer que l'effet est spécifique à la Tyr1, comme en déterminant la distribution des autres marques de phosphorylation du CTD.

#### 4.4 Conclusion

En conclusion, les résultats obtenus durant mes études de doctorat ont permis d'effectuer une mise à jour importante des différents modèles de terminaison transcriptionnelle chez la levure *S. cerevisiae*. Nous avons confirmé l'implication clé de la Ser2 et de la Thr4 du CTD dans la terminaison des gènes codants et nous avons démontré que la Ser5 n'est finalement pas nécessaire à la terminaison des gènes non-codants. Les résultats les plus significatifs sont toutefois ceux obtenus pour la Tyr1 où nous démontrons qu'elle n'est pas seulement nécessaire à la terminaison d'un nombre très limité de gènes codants, mais qu'elle est nécessaire à la presque totalité des gènes non-codants. De plus, la Tyr1 est impliquée dans la régulation d'évènements de pause de l'ARNPII dans la région 5' des gènes, aussi conservée chez les mammifères, qui sont nécessaires au NNS afin d'induire efficacement la terminaison. Finalement, les résultats obtenus pour le rôle de Glc7 dans la stabilité de Sen1 permettent enfin de fournir une explication *in vivo* du rôle de la phosphatase dans son action sur l'hélicase. Ces résultats pourraient aussi avoir des retombées intéressantes chez l'humain s'ils s'avèrent conservés.



## Bibliographie

1. Darwin, C., *The origin of species*. 1859.
2. Simunek, M., U. Hossfeld, and V. Wissemann, '*Rediscovery*' revised - the cooperation of Erich and Armin von Tschermak-Seysenegg in the context of the 'rediscovery' of Mendel's laws in 1899-1901. *Plant Biol (Stuttg)*, 2011. **13**(6): p. 835-41.
3. Satzinger, H., *Theodor and Marcella Boveri: chromosomes and cytoplasm in heredity and development*. *Nat Rev Genet*, 2008. **9**(3): p. 231-8.
4. Morgan, T.H., *Sex Limited Inheritance in Drosophila*. *Science*, 1910. **32**(812): p. 120-2.
5. Dahm, R., *Friedrich Miescher and the discovery of DNA*. *Dev Biol*, 2005. **278**(2): p. 274-88.
6. Griffith, F., *The Significance of Pneumococcal Types*. *J Hyg (Lond)*, 1928. **27**(2): p. 113-59.
7. Cobb, M., *Oswald Avery, DNA, and the transformation of biology*. *Curr Biol*, 2014. **24**(2): p. R55-60.
8. Hershey, A.D. and M. Chase, *Independent functions of viral protein and nucleic acid in growth of bacteriophage*. *J Gen Physiol*, 1952. **36**(1): p. 39-56.
9. Chargaff, E., B. Magasanik, and et al., *The nucleotide composition of ribonucleic acids*. *J Am Chem Soc*, 1949. **71**(4): p. 1513.
10. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. *Nature*, 1953. **171**(4356): p. 737-8.
11. Brenner, S., F. Jacob, and M. Meselson, *An unstable intermediate carrying information from genes to ribosomes for protein synthesis*. *Nature*, 1961. **190**: p. 576-581.
12. Gros, F., et al., *Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli*. *Nature*, 1961. **190**: p. 581-5.
13. Grunberg-Manago, M., P.J. Ortiz, and S. Ochoa, *Enzymic synthesis of polynucleotides. I. Polynucleotide phosphorylase of azotobacter vinelandii*. *Biochim Biophys Acta*, 1956. **20**(1): p. 269-85.
14. Bessman, M.J., et al., *Enzymic synthesis of deoxyribonucleic acid*. *Biochim Biophys Acta*, 1956. **21**(1): p. 197-8.

15. Roeder, R.G. and W.J. Rutter, *Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms*. Nature, 1969. **224**(5216): p. 234-7.
16. Kedinger, C., et al., *Alpha-amanitin: a specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus*. Biochem Biophys Res Commun, 1970. **38**(1): p. 165-71.
17. Werner, M., P. Thuriaux, and J. Soutourina, *Structure-function analysis of RNA polymerases I and III*. Curr Opin Struct Biol, 2009. **19**(6): p. 740-5.
18. Kwapisz, M., F. Beckouet, and P. Thuriaux, *Early evolution of eukaryotic DNA-dependent RNA polymerases*. Trends Genet, 2008. **24**(5): p. 211-5.
19. Young, R.A., *RNA polymerase II*. Annu Rev Biochem, 1991. **60**: p. 689-715.
20. Hahn, S., *Structure and mechanism of the RNA polymerase II transcription machinery*. Nat Struct Mol Biol, 2004. **11**(5): p. 394-403.
21. Orlicky, S.M., et al., *Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation*. J Biol Chem, 2001. **276**(13): p. 10097-102.
22. Rosenheck, S. and M. Choder, *Rpb4, a subunit of RNA polymerase II, enables the enzyme to transcribe at temperature extremes in vitro*. J Bacteriol, 1998. **180**(23): p. 6187-92.
23. Choder, M., *A growth rate-limiting process in the last growth phase of the yeast life cycle involves RPB4, a subunit of RNA polymerase II*. J Bacteriol, 1993. **175**(19): p. 6358-63.
24. Choder, M. and R.A. Young, *A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival*. Mol Cell Biol, 1993. **13**(11): p. 6984-91.
25. Li, S. and M.J. Smerdon, *Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in Saccharomyces cerevisiae*. EMBO J, 2002. **21**(21): p. 5921-9.
26. Farago, M., et al., *Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress*. Mol Biol Cell, 2003. **14**(7): p. 2744-55.
27. Lotan, R., et al., *The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs*. Genes Dev, 2005. **19**(24): p. 3004-16.

28. Harel-Sharvit, L., et al., *RNA polymerase II subunits link transcription and mRNA decay to translation*. Cell, 2010. **143**(4): p. 552-63.
29. Fu, J., et al., *Yeast RNA polymerase II at 5 Å resolution*. Cell, 1999. **98**(6): p. 799-810.
30. Cramer, P., et al., *Architecture of RNA polymerase II and implications for the transcription mechanism*. Science, 2000. **288**(5466): p. 640-9.
31. Woychik, N.A. and M. Hampsey, *The RNA polymerase II machinery: structure illuminates function*. Cell, 2002. **108**(4): p. 453-63.
32. Cramer, P., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution*. Science, 2001. **292**(5523): p. 1863-76.
33. Gnatt, A.L., et al., *Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution*. Science, 2001. **292**(5523): p. 1876-82.
34. Kolodziej, P.A., et al., *RNA polymerase II subunit composition, stoichiometry, and phosphorylation*. Mol Cell Biol, 1990. **10**(5): p. 1915-20.
35. Kayukawa, K., et al., *A serine residue in the N-terminal acidic region of rat RPB6, one of the common subunits of RNA polymerases, is exclusively phosphorylated by casein kinase II in vitro*. Gene, 1999. **234**(1): p. 139-47.
36. Armache, K.J., H. Kettenberger, and P. Cramer, *Architecture of initiation-competent 12-subunit RNA polymerase II*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6964-8.
37. Bushnell, D.A. and R.D. Kornberg, *Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6969-73.
38. Steitz, T.A., *A mechanism for all polymerases*. Nature, 1998. **391**(6664): p. 231-2.
39. Westover, K.D., D.A. Bushnell, and R.D. Kornberg, *Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center*. Cell, 2004. **119**(4): p. 481-9.
40. Sosunov, V., et al., *Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase*. EMBO J, 2003. **22**(9): p. 2234-44.
41. Svetlov, V. and E. Nudler, *Jamming the ratchet of transcription*. Nat Struct Mol Biol, 2008. **15**(8): p. 777-9.
42. Vassylyev, D.G., et al., *Structural basis for substrate loading in bacterial RNA polymerase*. Nature, 2007. **448**(7150): p. 163-8.

43. Kettenberger, H., K.J. Armache, and P. Cramer, *Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS*. Mol Cell, 2004. **16**(6): p. 955-65.
44. Brueckner, F., J. Ortiz, and P. Cramer, *A movie of the RNA polymerase nucleotide addition cycle*. Curr Opin Struct Biol, 2009. **19**(3): p. 294-9.
45. Kaplan, C.D., K.M. Larsson, and R.D. Kornberg, *The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin*. Mol Cell, 2008. **30**(5): p. 547-56.
46. Steitz, T.A., et al., *A unified polymerase mechanism for nonhomologous DNA and RNA polymerases*. Science, 1994. **266**(5193): p. 2022-5.
47. Abbondanzieri, E.A., et al., *Direct observation of base-pair stepping by RNA polymerase*. Nature, 2005. **438**(7067): p. 460-5.
48. Feig, M. and Z.F. Burton, *RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation*. Biophys J, 2010. **99**(8): p. 2577-86.
49. Brueckner, F. and P. Cramer, *Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation*. Nat Struct Mol Biol, 2008. **15**(8): p. 811-8.
50. Bushnell, D.A., P. Cramer, and R.D. Kornberg, *Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 Å resolution*. Proc Natl Acad Sci U S A, 2002. **99**(3): p. 1218-22.
51. Gong, X.Q., Y.A. Nedialkov, and Z.F. Burton, *Alpha-amanitin blocks translocation by human RNA polymerase II*. J Biol Chem, 2004. **279**(26): p. 27422-7.
52. Wang, D., et al., *Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis*. Cell, 2006. **127**(5): p. 941-54.
53. Hazelbaker, D.Z., et al., *Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination*. Mol Cell, 2013. **49**(1): p. 55-66.
54. Malik, I., et al., *Wide-ranging and unexpected consequences of altered Pol II catalytic activity in vivo*. Nucleic Acids Res, 2017. **45**(8): p. 4431-4451.
55. Kaplan, C.D., et al., *Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo*. PLoS Genet, 2012. **8**(4): p. e1002627.

56. Kireeva, M.L., et al., *Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase*. BMC Biophys, 2012. **5**: p. 11.
57. Kireeva, M.L., et al., *Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation*. Mol Cell, 2008. **30**(5): p. 557-66.
58. Malagon, F., et al., *Mutations in the Saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil*. Genetics, 2006. **172**(4): p. 2201-9.
59. Jimeno-Gonzalez, S., et al., *The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II*. Mol Cell, 2010. **37**(4): p. 580-7.
60. Corden, J.L., et al., *A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II*. Proc Natl Acad Sci U S A, 1985. **82**(23): p. 7934-8.
61. Allison, L.A., et al., *Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases*. Cell, 1985. **42**(2): p. 599-610.
62. Chapman, R.D., et al., *Molecular evolution of the RNA polymerase II CTD*. Trends Genet, 2008. **24**(6): p. 289-96.
63. Eick, D. and M. Geyer, *The RNA polymerase II carboxy-terminal domain (CTD) code*. Chem Rev, 2013. **113**(11): p. 8456-90.
64. Corden, J.L., *RNA polymerase II C-terminal domain: Tethering transcription to transcript and template*. Chem Rev, 2013. **113**(11): p. 8423-55.
65. Jeronimo, C., A.R. Bataille, and F. Robert, *The writers, readers, and functions of the RNA polymerase II C-terminal domain code*. Chem Rev, 2013. **113**(11): p. 8491-522.
66. Chapman, R.D., et al., *The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability*. Nucleic Acids Res, 2004. **32**(1): p. 35-44.
67. Chapman, R.D., M. Conrad, and D. Eick, *Role of the mammalian RNA polymerase II C-terminal domain (CTD) nonconsensus repeats in CTD stability and cell proliferation*. Mol Cell Biol, 2005. **25**(17): p. 7665-74.
68. Portz, B., et al., *Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain*. Nat Commun, 2017. **8**: p. 15231.
69. Yang, C. and J.W. Stiller, *Evolutionary diversity and taxon-specific modifications of the RNA polymerase II C-terminal domain*. Proc Natl Acad Sci U S A, 2014. **111**(16): p. 5920-5.

70. Harlen, K.M. and L.S. Churchman, *The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain*. Nat Rev Mol Cell Biol, 2017. **18**(4): p. 263-273.
71. Jeffreys, A.J., et al., *Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA*. Nature, 1988. **332**(6161): p. 278-81.
72. West, M.L. and J.L. Corden, *Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations*. Genetics, 1995. **140**(4): p. 1223-33.
73. Nonet, M., D. Sweetser, and R.A. Young, *Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II*. Cell, 1987. **50**(6): p. 909-15.
74. Allison, L.A., et al., *The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function*. Mol Cell Biol, 1988. **8**(1): p. 321-9.
75. Brickey, W.J. and A.L. Greenleaf, *Functional studies of the carboxy-terminal repeat domain of Drosophila RNA polymerase II in vivo*. Genetics, 1995. **140**(2): p. 599-613.
76. Bartolomei, M.S., et al., *Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II*. Mol Cell Biol, 1988. **8**(1): p. 330-9.
77. Zehring, W.A., et al., *The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro*. Proc Natl Acad Sci U S A, 1988. **85**(11): p. 3698-702.
78. Schneider, S., et al., *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, 2010. **30**(10): p. 2353-64.
79. Coudreuse, D., et al., *A gene-specific requirement of RNA polymerase II CTD phosphorylation for sexual differentiation in S. pombe*. Curr Biol, 2010. **20**(12): p. 1053-64.
80. Zhang, D.W., et al., *Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination*. J Biol Chem, 2012. **287**(11): p. 8541-51.
81. Stiller, J.W., B.L. McConaughy, and B.D. Hall, *Evolutionary complementation for polymerase II CTD function*. Yeast, 2000. **16**(1): p. 57-64.

82. Schwer, B. and S. Shuman, *Deciphering the RNA polymerase II CTD code in fission yeast*. Mol Cell, 2011. **43**(2): p. 311-8.
83. Schwer, B., A.M. Sanchez, and S. Shuman, *Punctuation and syntax of the RNA polymerase II CTD code in fission yeast*. Proc Natl Acad Sci U S A, 2012. **109**(44): p. 18024-9.
84. Inada, M., et al., *Phospho-site mutants of the RNA Polymerase II C-terminal domain alter subtelomeric gene expression and chromatin modification state in fission yeast*. Nucleic Acids Res, 2016. **44**(19): p. 9180-9189.
85. Hintermair, C., et al., *Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation*. EMBO J, 2012. **31**(12): p. 2784-97.
86. Hsin, J.P., A. Sheth, and J.L. Manley, *RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing*. Science, 2011. **334**(6056): p. 683-6.
87. Descostes, N., et al., *Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells*. Elife, 2014. **3**: p. e02105.
88. Shah, N., et al., *Tyrosine-1 of RNA Polymerase II CTD Controls Global Termination of Gene Transcription in Mammals*. Mol Cell, 2018. **69**(1): p. 48-61 e6.
89. Liu, P., et al., *Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain*. Mol Biol Evol, 2010. **27**(11): p. 2628-41.
90. Stiller, J.W. and M.S. Cook, *Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs*. Eukaryot Cell, 2004. **3**(3): p. 735-40.
91. Suzuki, M., *SPXX, a frequent sequence motif in gene regulatory proteins*. J Mol Biol, 1989. **207**(1): p. 61-84.
92. Harding, M.M., *NMR studies on YSPTSPSY: implications for the design of DNA bisintercalators*. J Med Chem, 1992. **35**(25): p. 4658-64.
93. Cagas, P.M. and J.L. Corden, *Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II*. Proteins, 1995. **21**(2): p. 149-60.
94. Noble, C.G., et al., *Key features of the interaction between Pcf11 CID and RNA polymerase II CTD*. Nat Struct Mol Biol, 2005. **12**(2): p. 144-51.

95. Bienkiewicz, E.A., A. Moon Woody, and R.W. Woody, *Conformation of the RNA polymerase II C-terminal domain: circular dichroism of long and short fragments*. J Mol Biol, 2000. **297**(1): p. 119-33.
96. Meredith, G.D., et al., *The C-terminal domain revealed in the structure of RNA polymerase II*. J Mol Biol, 1996. **258**(3): p. 413-9.
97. Meinhart, A., et al., *A structural perspective of CTD function*. Genes Dev, 2005. **19**(12): p. 1401-15.
98. Meinhart, A. and P. Cramer, *Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors*. Nature, 2004. **430**(6996): p. 223-6.
99. Zhang, J. and J.L. Corden, *Phosphorylation causes a conformational change in the carboxyl-terminal domain of the mouse RNA polymerase II largest subunit*. J Biol Chem, 1991. **266**(4): p. 2297-302.
100. Laybourn, P.J. and M.E. Dahmus, *Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit IIa/o*. J Biol Chem, 1989. **264**(12): p. 6693-8.
101. Kwon, I., et al., *Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains*. Cell, 2013. **155**(5): p. 1049-1060.
102. Burke, K.A., et al., *Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II*. Mol Cell, 2015. **60**(2): p. 231-41.
103. Boehning, M., et al., *RNA polymerase II clustering through carboxy-terminal domain phase separation*. Nat Struct Mol Biol, 2018. **25**(9): p. 833-840.
104. Lu, H., et al., *Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II*. Nature, 2018. **558**(7709): p. 318-323.
105. Cisse, II, et al., *Real-time dynamics of RNA polymerase II clustering in live human cells*. Science, 2013. **341**(6146): p. 664-7.
106. Cho, W.K., et al., *RNA Polymerase II cluster dynamics predict mRNA output in living cells*. Elife, 2016. **5**.
107. Boeynaems, S., et al., *Protein Phase Separation: A New Phase in Cell Biology*. Trends Cell Biol, 2018. **28**(6): p. 420-435.



108. Cadena, D.L. and M.E. Dahmus, *Messenger RNA synthesis in mammalian cells is catalyzed by the phosphorylated form of RNA polymerase II*. J Biol Chem, 1987. **262**(26): p. 12468-74.
109. Bartholomew, B., M.E. Dahmus, and C.F. Meares, *RNA contacts subunits Ilo and Iic in HeLa RNA polymerase II transcription complexes*. J Biol Chem, 1986. **261**(30): p. 14226-31.
110. Lu, H., et al., *The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex*. Proc Natl Acad Sci U S A, 1991. **88**(22): p. 10004-8.
111. Laybourn, P.J. and M.E. Dahmus, *Phosphorylation of RNA polymerase IIA occurs subsequent to interaction with the promoter and before the initiation of transcription*. J Biol Chem, 1990. **265**(22): p. 13165-73.
112. Chesnut, J.D., J.H. Stephens, and M.E. Dahmus, *The interaction of RNA polymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit Ila*. J Biol Chem, 1992. **267**(15): p. 10500-6.
113. Kang, M.E. and M.E. Dahmus, *RNA polymerases IIA and IIO have distinct roles during transcription from the TATA-less murine dihydrofolate reductase promoter*. J Biol Chem, 1993. **268**(33): p. 25033-40.
114. Robinson, P.J., et al., *Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex*. Cell, 2016. **166**(6): p. 1411-1422 e16.
115. Robinson, P.J., et al., *Structure of the mediator head module bound to the carboxy-terminal domain of RNA polymerase II*. Proc Natl Acad Sci U S A, 2012. **109**(44): p. 17931-5.
116. Myers, L.C., et al., *The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain*. Genes Dev, 1998. **12**(1): p. 45-54.
117. Yuryev, A. and J.L. Corden, *Suppression analysis reveals a functional difference between the serines in positions two and five in the consensus sequence of the C-terminal domain of yeast RNA polymerase II*. Genetics, 1996. **143**(2): p. 661-71.
118. Zhang, J. and J.L. Corden, *Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit*. J Biol Chem, 1991. **266**(4): p. 2290-6.

119. Lee, J.M. and A.L. Greenleaf, *A protein kinase that phosphorylates the C-terminal repeat domain of the largest subunit of RNA polymerase II*. Proc Natl Acad Sci U S A, 1989. **86**(10): p. 3624-8.
120. Weeks, J.R., et al., *Locus-specific variation in phosphorylation state of RNA polymerase II in vivo: correlations with gene activity and transcript processing*. Genes Dev, 1993. **7**(12A): p. 2329-44.
121. O'Brien, T., et al., *Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation*. Nature, 1994. **370**(6484): p. 75-7.
122. Patturajan, M., et al., *Growth-related changes in phosphorylation of yeast RNA polymerase II*. J Biol Chem, 1998. **273**(8): p. 4689-94.
123. Bregman, D.B., et al., *Cytostellin distributes to nuclear regions enriched with splicing factors*. J Cell Sci, 1994. **107 ( Pt 3)**: p. 387-96.
124. Warren, S.L., et al., *Cytostellin: a novel, highly conserved protein that undergoes continuous redistribution during the cell cycle*. J Cell Sci, 1992. **103 ( Pt 2)**: p. 381-8.
125. Komarnitsky, P., E.J. Cho, and S. Buratowski, *Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription*. Genes Dev, 2000. **14**(19): p. 2452-60.
126. Buratowski, S., *The CTD code*. Nat Struct Biol, 2003. **10**(9): p. 679-80.
127. Chapman, R.D., et al., *Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7*. Science, 2007. **318**(5857): p. 1780-2.
128. Mayer, A., et al., *CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II*. Science, 2012. **336**(6089): p. 1723-5.
129. Kim, H., et al., *Gene-specific RNA polymerase II phosphorylation and the CTD code*. Nat Struct Mol Biol, 2010. **17**(10): p. 1279-86.
130. 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, 2012. **45**(2): p. 158-70.
131. Mayer, A., et al., *Uniform transitions of the general RNA polymerase II transcription complex*. Nat Struct Mol Biol, 2010. **17**(10): p. 1272-8.
132. Tietjen, J.R., et al., *Chemical-genomic dissection of the CTD code*. Nat Struct Mol Biol, 2010. **17**(9): p. 1154-61.

133. Milligan, L., et al., *Strand-specific, high-resolution mapping of modified RNA polymerase II*. Mol Syst Biol, 2016. **12**(6): p. 874.
134. Harlen, K.M., et al., *Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue*. Cell Rep, 2016. **15**(10): p. 2147-2158.
135. Koch, F., et al., *Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters*. Nat Struct Mol Biol, 2011. **18**(8): p. 956-63.
136. Hsin, J.P., et al., *RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells*. Elife, 2014. **3**: p. e02112.
137. Larochelle, M., et al., *Common mechanism of transcription termination at coding and noncoding RNA genes in fission yeast*. Nat Commun, 2018. **9**(1): p. 4364.
138. Akhtar, M.S., et al., *TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II*. Mol Cell, 2009. **34**(3): p. 387-93.
139. Jeronimo, C., P. Collin, and F. Robert, *The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain*. J Mol Biol, 2016. **428**(12): p. 2607-2622.
140. Suh, H., et al., *Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II*. Mol Cell, 2016. **61**(2): p. 297-304.
141. Schuller, R., et al., *Heptad-Specific Phosphorylation of RNA Polymerase II CTD*. Mol Cell, 2016. **61**(2): p. 305-14.
142. Schrieck, A., et al., *RNA polymerase II termination involves C-terminal-domain tyrosine dephosphorylation by CPF subunit Glc7*. Nat Struct Mol Biol, 2014. **21**(2): p. 175-179.
143. Feaver, W.J., et al., *CTD kinase associated with yeast RNA polymerase II initiation factor b*. Cell, 1991. **67**(6): p. 1223-30.
144. Lu, H., et al., *Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II*. Nature, 1992. **358**(6388): p. 641-5.
145. Glover-Cutter, K., et al., *TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II*. Mol Cell Biol, 2009. **29**(20): p. 5455-64.

146. Jeronimo, C. and F. Robert, *Kin28 regulates the transient association of Mediator with core promoters*. Nat Struct Mol Biol, 2014. **21**(5): p. 449-55.
147. Wong, K.H., Y. Jin, and K. Struhl, *TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape*. Mol Cell, 2014. **54**(4): p. 601-12.
148. Akoulitchev, S., et al., *Requirement for TFIIH kinase activity in transcription by RNA polymerase II*. Nature, 1995. **377**(6549): p. 557-60.
149. Liu, Y., et al., *Two cyclin-dependent kinases promote RNA polymerase II transcription and formation of the scaffold complex*. Mol Cell Biol, 2004. **24**(4): p. 1721-35.
150. Svejstrup, J.Q., et al., *Evidence for a mediator cycle at the initiation of transcription*. Proc Natl Acad Sci U S A, 1997. **94**(12): p. 6075-8.
151. Sogaard, T.M. and J.Q. Svejstrup, *Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator*. J Biol Chem, 2007. **282**(19): p. 14113-20.
152. Muse, G.W., et al., *RNA polymerase is poised for activation across the genome*. Nat Genet, 2007. **39**(12): p. 1507-11.
153. Lee, C., et al., *NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila*. Mol Cell Biol, 2008. **28**(10): p. 3290-300.
154. Zeitlinger, J., et al., *RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo*. Nat Genet, 2007. **39**(12): p. 1512-6.
155. Guenther, M.G., et al., *A chromatin landmark and transcription initiation at most promoters in human cells*. Cell, 2007. **130**(1): p. 77-88.
156. Kim, T.H., et al., *A high-resolution map of active promoters in the human genome*. Nature, 2005. **436**(7052): p. 876-80.
157. Yamaguchi, Y., et al., *NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation*. Cell, 1999. **97**(1): p. 41-51.
158. Wada, T., et al., *DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs*. Genes Dev, 1998. **12**(3): p. 343-56.

159. Yamaguchi, Y., H. Shibata, and H. Handa, *Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond*. *Biochim Biophys Acta*, 2013. **1829**(1): p. 98-104.
160. Narita, T., et al., *Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex*. *Mol Cell Biol*, 2003. **23**(6): p. 1863-73.
161. Marshall, N.F. and D.H. Price, *Purification of P-TEFb, a transcription factor required for the transition into productive elongation*. *J Biol Chem*, 1995. **270**(21): p. 12335-8.
162. Peng, J., et al., *Identification of multiple cyclin subunits of human P-TEFb*. *Genes Dev*, 1998. **12**(5): p. 755-62.
163. Peng, J., N.F. Marshall, and D.H. Price, *Identification of a cyclin subunit required for the function of Drosophila P-TEFb*. *J Biol Chem*, 1998. **273**(22): p. 13855-60.
164. Murray, S., et al., *Phosphorylation of the RNA polymerase II carboxy-terminal domain by the Bur1 cyclin-dependent kinase*. *Mol Cell Biol*, 2001. **21**(13): p. 4089-96.
165. Yamada, T., et al., *P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation*. *Mol Cell*, 2006. **21**(2): p. 227-37.
166. Fujinaga, K., et al., *Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element*. *Mol Cell Biol*, 2004. **24**(2): p. 787-95.
167. Czudnochowski, N., C.A. Bosken, and M. Geyer, *Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition*. *Nat Commun*, 2012. **3**: p. 842.
168. Ramanathan, Y., et al., *Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences*. *J Biol Chem*, 2001. **276**(14): p. 10913-20.
169. Bartkowiak, B., et al., *CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1*. *Genes Dev*, 2010. **24**(20): p. 2303-16.
170. Sterner, D.E., et al., *The yeast carboxyl-terminal repeat domain kinase CTDK-I is a divergent cyclin-cyclin-dependent kinase complex*. *Mol Cell Biol*, 1995. **15**(10): p. 5716-24.

171. Lee, J.M. and A.L. Greenleaf, *CTD kinase large subunit is encoded by CTK1, a gene required for normal growth of Saccharomyces cerevisiae*. *Gene Expr*, 1991. **1**(2): p. 149-67.
172. Qiu, H., C. Hu, and A.G. Hinnebusch, *Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters*. *Mol Cell*, 2009. **33**(6): p. 752-62.
173. Cho, E.J., et al., *Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain*. *Genes Dev*, 2001. **15**(24): p. 3319-29.
174. Keogh, M.C., V. Podolny, and S. Buratowski, *Bur1 kinase is required for efficient transcription elongation by RNA polymerase II*. *Mol Cell Biol*, 2003. **23**(19): p. 7005-18.
175. Jones, J.C., et al., *C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats*. *J Biol Chem*, 2004. **279**(24): p. 24957-64.
176. Viladevall, L., et al., *TFIIF and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast*. *Mol Cell*, 2009. **33**(6): p. 738-51.
177. Larochelle, S., et al., *Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II*. *Nat Struct Mol Biol*, 2012. **19**(11): p. 1108-15.
178. Bosken, C.A., et al., *The structure and substrate specificity of human Cdk12/Cyclin K*. *Nat Commun*, 2014. **5**: p. 3505.
179. Liang, K., et al., *Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing*. *Mol Cell Biol*, 2015. **35**(6): p. 928-38.
180. Blazek, D., et al., *The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes*. *Genes Dev*, 2011. **25**(20): p. 2158-72.
181. Bartkowiak, B., C. Yan, and A.L. Greenleaf, *Engineering an analog-sensitive CDK12 cell line using CRISPR/Cas*. *Biochim Biophys Acta*, 2015. **1849**(9): p. 1179-87.
182. Devaiah, B.N., et al., *BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain*. *Proc Natl Acad Sci U S A*, 2012. **109**(18): p. 6927-32.

183. Devaiah, B.N. and D.S. Singer, *Cross-talk among RNA polymerase II kinases modulates C-terminal domain phosphorylation*. J Biol Chem, 2012. **287**(46): p. 38755-66.
184. Yang, Z., et al., *Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4*. Mol Cell, 2005. **19**(4): p. 535-45.
185. Jang, M.K., et al., *The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription*. Mol Cell, 2005. **19**(4): p. 523-34.
186. Zhang, W., et al., *Bromodomain-containing protein 4 (BRD4) regulates RNA polymerase II serine 2 phosphorylation in human CD4+ T cells*. J Biol Chem, 2012. **287**(51): p. 43137-55.
187. Baskaran, R., M.E. Dahmus, and J.Y. Wang, *Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain*. Proc Natl Acad Sci U S A, 1993. **90**(23): p. 11167-71.
188. Baskaran, R., et al., *Ataxia telangiectasia mutant protein activates c-Abl tyrosine kinase in response to ionizing radiation*. Nature, 1997. **387**(6632): p. 516-9.
189. Baskaran, R., G.G. Chiang, and J.Y. Wang, *Identification of a binding site in c-Abl tyrosine kinase for the C-terminal repeated domain of RNA polymerase II*. Mol Cell Biol, 1996. **16**(7): p. 3361-9.
190. Baskaran, R., S.R. Escobar, and J.Y. Wang, *Nuclear c-Abl is a COOH-terminal repeated domain (CTD)-tyrosine (CTD)-tyrosine kinase-specific for the mammalian RNA polymerase II: possible role in transcription elongation*. Cell Growth Differ, 1999. **10**(6): p. 387-96.
191. Baskaran, R., et al., *Tyrosine phosphorylation of RNA polymerase II carboxyl-terminal domain by the Abl-related gene product*. J Biol Chem, 1997. **272**(30): p. 18905-9.
192. Yurko, N., et al., *MPK1/SLT2 Links Multiple Stress Responses with Gene Expression in Budding Yeast by Phosphorylating Tyr1 of the RNAP II CTD*. Mol Cell, 2017. **68**(5): p. 913-925 e3.
193. Mosley, A.L., et al., *Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation*. Mol Cell, 2009. **34**(2): p. 168-78.

194. Egloff, S., et al., *Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes*. Mol Cell, 2012. **45**(1): p. 111-22.
195. Ni, Z., et al., *RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation*. Nat Struct Mol Biol, 2014. **21**(8): p. 686-695.
196. Smith-Kinnaman, W.R., et al., *The interactome of the atypical phosphatase Rtr1 in Saccharomyces cerevisiae*. Mol Biosyst, 2014. **10**(7): p. 1730-41.
197. Cho, H., et al., *A protein phosphatase functions to recycle RNA polymerase II*. Genes Dev, 1999. **13**(12): p. 1540-52.
198. Kobor, M.S., et al., *An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae*. Mol Cell, 1999. **4**(1): p. 55-62.
199. Kong, S.E., et al., *Interaction of Fcp1 phosphatase with elongating RNA polymerase II holoenzyme, enzymatic mechanism of action, and genetic interaction with elongator*. J Biol Chem, 2005. **280**(6): p. 4299-306.
200. Hausmann, S. and S. Shuman, *Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5*. J Biol Chem, 2002. **277**(24): p. 21213-20.
201. Hausmann, S., H. Erdjument-Bromage, and S. Shuman, *Schizosaccharomyces pombe carboxyl-terminal domain (CTD) phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping*. J Biol Chem, 2004. **279**(12): p. 10892-900.
202. Nedeá, E., et al., *Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends*. J Biol Chem, 2003. **278**(35): p. 33000-10.
203. He, X., et al., *Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1*. Genes Dev, 2003. **17**(8): p. 1030-42.
204. Krishnamurthy, S., et al., *Ssu72 Is an RNA polymerase II CTD phosphatase*. Mol Cell, 2004. **14**(3): p. 387-94.
205. Hausmann, S., et al., *Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases*. J Biol Chem, 2005. **280**(45): p. 37681-8.



206. Xiang, K., et al., *Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex*. Nature, 2010. **467**(7316): p. 729-33.
207. Xiang, K., J.L. Manley, and L. Tong, *An unexpected binding mode for a Pol II CTD peptide phosphorylated at Ser7 in the active site of the CTD phosphatase Ssu72*. Genes Dev, 2012. **26**(20): p. 2265-70.
208. Werner-Allen, J.W., et al., *cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72*. J Biol Chem, 2011. **286**(7): p. 5717-26.
209. Krishnamurthy, S., et al., *Functional interaction of the Ess1 prolyl isomerase with components of the RNA polymerase II initiation and termination machineries*. Mol Cell Biol, 2009. **29**(11): p. 2925-34.
210. Kecman, T., et al., *Elongation/Termination Factor Exchange Mediated by PPI Phosphatase Orchestrates Transcription Termination*. Cell Rep, 2018. **25**(1): p. 259-269 e5.
211. Ranuncolo, S.M., et al., *Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo*. J Biol Chem, 2012. **287**(28): p. 23549-61.
212. Kelly, W.G., M.E. Dahmus, and G.W. Hart, *RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc*. J Biol Chem, 1993. **268**(14): p. 10416-24.
213. Comer, F.I. and G.W. Hart, *Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II*. Biochemistry, 2001. **40**(26): p. 7845-52.
214. Sims, R.J., 3rd, et al., *The C-terminal domain of RNA polymerase II is modified by site-specific methylation*. Science, 2011. **332**(6025): p. 99-103.
215. Zhao, D.Y., et al., *SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination*. Nature, 2016. **529**(7584): p. 48-53.
216. Voss, K., et al., *Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II*. Transcription, 2015. **6**(5): p. 91-101.
217. Dias, J.D., et al., *Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells*. Elife, 2015. **4**.

218. Schroder, S., et al., *Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells*. Mol Cell, 2013. **52**(3): p. 314-24.
219. Li, H., et al., *Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells*. Mol Cell Biol, 2007. **27**(15): p. 5296-305.
220. Proudfoot, N.J., A. Furger, and M.J. Dye, *Integrating mRNA processing with transcription*. Cell, 2002. **108**(4): p. 501-12.
221. Tsai, K.L., et al., *A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction*. Nat Struct Mol Biol, 2013. **20**(5): p. 611-9.
222. Payne, J.M., P.J. Laybourn, and M.E. Dahmus, *The transition of RNA polymerase II from initiation to elongation is associated with phosphorylation of the carboxyl-terminal domain of subunit IIa*. J Biol Chem, 1989. **264**(33): p. 19621-9.
223. Tsai, K.L., et al., *Mediator structure and rearrangements required for holoenzyme formation*. Nature, 2017. **544**(7649): p. 196-201.
224. Hagler, J. and S. Shuman, *A freeze-frame view of eukaryotic transcription during elongation and capping of nascent mRNA*. Science, 1992. **255**(5047): p. 983-6.
225. Ghosh, A. and C.D. Lima, *Enzymology of RNA cap synthesis*. Wiley Interdiscip Rev RNA, 2010. **1**(1): p. 152-72.
226. Cho, E.J., et al., *mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain*. Genes Dev, 1997. **11**(24): p. 3319-26.
227. McCracken, S., et al., *5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II*. Genes Dev, 1997. **11**(24): p. 3306-18.
228. Yue, Z., et al., *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, 1997. **94**(24): p. 12898-903.
229. Ho, C.K., et al., *The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II*. J Biol Chem, 1998. **273**(16): p. 9577-85.

230. Ho, C.K. and S. Shuman, *Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme*. Mol Cell, 1999. **3**(3): p. 405-11.
231. Fabrega, C., et al., *Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II*. Mol Cell, 2003. **11**(6): p. 1549-61.
232. Doamekpor, S.K., et al., *How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes*. Genes Dev, 2014. **28**(12): p. 1323-36.
233. Ghosh, A., S. Shuman, and C.D. Lima, *Structural insights to how mammalian capping enzyme reads the CTD code*. Mol Cell, 2011. **43**(2): p. 299-310.
234. Kim, M., et al., *The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II*. Nature, 2004. **432**(7016): p. 517-22.
235. Lunde, B.M., et al., *Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain*. Nat Struct Mol Biol, 2010. **17**(10): p. 1195-201.
236. Steinmetz, E.J., et al., *RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts*. Nature, 2001. **413**(6853): p. 327-31.
237. Steinmetz, E.J. and D.A. Brow, *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, 1996. **16**(12): p. 6993-7003.
238. Conrad, N.K., et al., *A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II*. Genetics, 2000. **154**(2): p. 557-71.
239. Barilla, D., B.A. Lee, and N.J. Proudfoot, *Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A, 2001. **98**(2): p. 445-50.
240. Licatalosi, D.D., et al., *Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II*. Mol Cell, 2002. **9**(5): p. 1101-11.
241. Vasiljeva, L., et al., *The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain*. Nat Struct Mol Biol, 2008. **15**(8): p. 795-804.

242. Becker, R., B. Loll, and A. Meinhart, *Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II*. J Biol Chem, 2008. **283**(33): p. 22659-69.
243. Patturajan, M., et al., *A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II*. Mol Cell Biol, 1998. **18**(4): p. 2406-15.
244. Ni, Z., et al., *Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B*. Transcription, 2011. **2**(5): p. 237-42.
245. Kubicek, K., et al., *Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1*. Genes Dev, 2012. **26**(17): p. 1891-6.
246. Jasnovidova, O., et al., *Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p*. EMBO Rep, 2017. **18**(6): p. 906-913.
247. Nemecek, C.M., et al., *Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner*. Proc Natl Acad Sci U S A, 2017. **114**(20): p. E3944-E3953.
248. Sadowski, M., et al., *Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination*. EMBO J, 2003. **22**(9): p. 2167-77.
249. Liu, B.A. and P.D. Nash, *Evolution of SH2 domains and phosphotyrosine signalling networks*. Philos Trans R Soc Lond B Biol Sci, 2012. **367**(1602): p. 2556-73.
250. Dengl, S., et al., *Structure and in vivo requirement of the yeast Spt6 SH2 domain*. J Mol Biol, 2009. **389**(1): p. 211-25.
251. Burugula, B.B., et al., *Histone deacetylases and phosphorylated polymerase II C-terminal domain recruit Spt6 for cotranscriptional histone reassembly*. Mol Cell Biol, 2014. **34**(22): p. 4115-29.
252. Liu, J., et al., *Solution structure of tandem SH2 domains from Spt6 protein and their binding to the phosphorylated RNA polymerase II C-terminal domain*. J Biol Chem, 2011. **286**(33): p. 29218-26.
253. Vos, S.M., et al., *Structure of activated transcription complex Pol II-DSIF-PAF-SPT6*. Nature, 2018. **560**(7720): p. 607-612.

254. Sdano, M.A., et al., *A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription*. *Elife*, 2017. **6**.
255. Kizer, K.O., et al., *A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation*. *Mol Cell Biol*, 2005. **25**(8): p. 3305-16.
256. Vojnic, E., et al., *Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription*. *J Biol Chem*, 2006. **281**(1): p. 13-5.
257. Bork, P. and M. Sudol, *The WW domain: a signalling site in dystrophin?* *Trends Biochem Sci*, 1994. **19**(12): p. 531-3.
258. Sudol, M., et al., *Characterization of a novel protein-binding module--the WW domain*. *FEBS Lett*, 1995. **369**(1): p. 67-71.
259. Chen, H.I. and M. Sudol, *The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules*. *Proc Natl Acad Sci U S A*, 1995. **92**(17): p. 7819-23.
260. Sudol, M., *Structure and function of the WW domain*. *Prog Biophys Mol Biol*, 1996. **65**(1-2): p. 113-32.
261. Gemmill, T.R., X. Wu, and S.D. Hanes, *Vanishingly low levels of Ess1 prolyl-isomerase activity are sufficient for growth in Saccharomyces cerevisiae*. *J Biol Chem*, 2005. **280**(16): p. 15510-7.
262. Ma, Z., et al., *Multiple roles for the Ess1 prolyl isomerase in the RNA polymerase II transcription cycle*. *Mol Cell Biol*, 2012. **32**(17): p. 3594-607.
263. Li, Z., et al., *The structure of the Candida albicans Ess1 prolyl isomerase reveals a well-ordered linker that restricts domain mobility*. *Biochemistry*, 2005. **44**(16): p. 6180-9.
264. Ranganathan, R., et al., *Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent*. *Cell*, 1997. **89**(6): p. 875-86.
265. Verdecia, M.A., et al., *Structural basis for phosphoserine-proline recognition by group IV WW domains*. *Nat Struct Biol*, 2000. **7**(8): p. 639-43.

266. Bedford, M.T. and P. Leder, *The FF domain: a novel motif that often accompanies WW domains*. Trends Biochem Sci, 1999. **24**(7): p. 264-5.
267. Morris, D.P. and A.L. Greenleaf, *The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II*. J Biol Chem, 2000. **275**(51): p. 39935-43.
268. Carty, S.M., et al., *Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II*. Proc Natl Acad Sci U S A, 2000. **97**(16): p. 9015-20.
269. MacKellar, A.L. and A.L. Greenleaf, *Cotranscriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II*. J Biol Chem, 2011. **286**(42): p. 36385-95.
270. Sikorsky, T., et al., *Recognition of asymmetrically dimethylated arginine by TDRD3*. Nucleic Acids Res, 2012. **40**(22): p. 11748-55.
271. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription*. Cell, 2007. **128**(4): p. 707-19.
272. Cairns, B.R., *The logic of chromatin architecture and remodelling at promoters*. Nature, 2009. **461**(7261): p. 193-8.
273. Jiang, C. and B.F. Pugh, *Nucleosome positioning and gene regulation: advances through genomics*. Nat Rev Genet, 2009. **10**(3): p. 161-72.
274. Smale, S.T. and J.T. Kadonaga, *The RNA polymerase II core promoter*. Annu Rev Biochem, 2003. **72**: p. 449-79.
275. Juven-Gershon, T. and J.T. Kadonaga, *Regulation of gene expression via the core promoter and the basal transcriptional machinery*. Dev Biol, 2010. **339**(2): p. 225-9.
276. Yang, C., et al., *Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters*. Gene, 2007. **389**(1): p. 52-65.
277. Rhee, H.S. and B.F. Pugh, *Genome-wide structure and organization of eukaryotic pre-initiation complexes*. Nature, 2012. **483**(7389): p. 295-301.
278. Hampsey, M., *Molecular genetics of the RNA polymerase II general transcriptional machinery*. Microbiol Mol Biol Rev, 1998. **62**(2): p. 465-503.

279. Zhang, Z. and F.S. Dietrich, *Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE*. Nucleic Acids Res, 2005. **33**(9): p. 2838-51.
280. Rando, O.J. and F. Winston, *Chromatin and transcription in yeast*. Genetics, 2012. **190**(2): p. 351-87.
281. Huisinga, K.L. and B.F. Pugh, *A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae*. Mol Cell, 2004. **13**(4): p. 573-85.
282. Basehoar, A.D., S.J. Zanton, and B.F. Pugh, *Identification and distinct regulation of yeast TATA box-containing genes*. Cell, 2004. **116**(5): p. 699-709.
283. Baptista, T., et al., *SAGA Is a General Cofactor for RNA Polymerase II Transcription*. Mol Cell, 2017. **68**(1): p. 130-143 e5.
284. Warfield, L., et al., *Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID*. Mol Cell, 2017. **68**(1): p. 118-129 e5.
285. Grant, P.A., et al., *Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex*. Genes Dev, 1997. **11**(13): p. 1640-50.
286. Ricci, A.R., J. Genereaux, and C.J. Brandl, *Components of the SAGA histone acetyltransferase complex are required for repressed transcription of ARG1 in rich medium*. Mol Cell Biol, 2002. **22**(12): p. 4033-42.
287. Helmlinger, D., et al., *The S. pombe SAGA complex controls the switch from proliferation to sexual differentiation through the opposing roles of its subunits Gcn5 and Spt8*. Genes Dev, 2008. **22**(22): p. 3184-95.
288. Rodriguez-Navarro, S., *Insights into SAGA function during gene expression*. EMBO Rep, 2009. **10**(8): p. 843-50.
289. Chen, X.F., et al., *Mediator and SAGA have distinct roles in Pol II preinitiation complex assembly and function*. Cell Rep, 2012. **2**(5): p. 1061-7.
290. Hahn, S. and E.T. Young, *Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators*. Genetics, 2011. **189**(3): p. 705-36.

291. Sterner, D.E., et al., *Functional organization of the yeast SAGA complex: distinct components involved in structural integrity, nucleosome acetylation, and TATA-binding protein interaction*. Mol Cell Biol, 1999. **19**(1): p. 86-98.
292. Grant, P.A., et al., *A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation*. Cell, 1998. **94**(1): p. 45-53.
293. Wu, P.Y., et al., *Molecular architecture of the S. cerevisiae SAGA complex*. Mol Cell, 2004. **15**(2): p. 199-208.
294. Marcus, G.A., et al., *Functional similarity and physical association between GCN5 and ADA2: putative transcriptional adaptors*. EMBO J, 1994. **13**(20): p. 4807-15.
295. Horiuchi, J., et al., *ADA3, a putative transcriptional adaptor, consists of two separable domains and interacts with ADA2 and GCN5 in a trimeric complex*. Mol Cell Biol, 1995. **15**(3): p. 1203-9.
296. Balasubramanian, R., et al., *Role of the Ada2 and Ada3 transcriptional coactivators in histone acetylation*. J Biol Chem, 2002. **277**(10): p. 7989-95.
297. Henry, K.W., et al., *Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8*. Genes Dev, 2003. **17**(21): p. 2648-63.
298. Daniel, J.A., et al., *Deubiquitination of histone H2B by a yeast acetyltransferase complex regulates transcription*. J Biol Chem, 2004. **279**(3): p. 1867-71.
299. Rodriguez-Navarro, S., et al., *Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery*. Cell, 2004. **116**(1): p. 75-86.
300. Lee, K.K., et al., *The deubiquitylation activity of Ubp8 is dependent upon Sgf11 and its association with the SAGA complex*. Mol Cell Biol, 2005. **25**(3): p. 1173-82.
301. Shukla, A., P. Bajwa, and S.R. Bhaumik, *SAGA-associated Sgf73p facilitates formation of the preinitiation complex assembly at the promoters either in a HAT-dependent or independent manner in vivo*. Nucleic Acids Res, 2006. **34**(21): p. 6225-32.



302. Wyce, A., et al., *H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex*. Mol Cell, 2007. **27**(2): p. 275-88.
303. Grant, P.A., et al., *The ATM-related cofactor Tra1 is a component of the purified SAGA complex*. Mol Cell, 1998. **2**(6): p. 863-7.
304. Brown, C.E., et al., *Recruitment of HAT complexes by direct activator interactions with the ATM-related Tra1 subunit*. Science, 2001. **292**(5525): p. 2333-7.
305. Bhaumik, S.R., et al., *In vivo target of a transcriptional activator revealed by fluorescence resonance energy transfer*. Genes Dev, 2004. **18**(3): p. 333-43.
306. Mohibullah, N. and S. Hahn, *Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3*. Genes Dev, 2008. **22**(21): p. 2994-3006.
307. Green, M.R., *TBP-associated factors (TAFII)s: multiple, selective transcriptional mediators in common complexes*. Trends Biochem Sci, 2000. **25**(2): p. 59-63.
308. Muller, F., A. Zaucker, and L. Tora, *Developmental regulation of transcription initiation: more than just changing the actors*. Curr Opin Genet Dev, 2010. **20**(5): p. 533-40.
309. Maston, G.A., et al., *Non-canonical TAF complexes regulate active promoters in human embryonic stem cells*. Elife, 2012. **1**: p. e00068.
310. Cler, E., et al., *Recent advances in understanding the structure and function of general transcription factor TFIID*. Cell Mol Life Sci, 2009. **66**(13): p. 2123-34.
311. Blair, R.H., J.A. Goodrich, and J.F. Kugel, *Single-molecule fluorescence resonance energy transfer shows uniformity in TATA binding protein-induced DNA bending and heterogeneity in bending kinetics*. Biochemistry, 2012. **51**(38): p. 7444-55.
312. Kim, J.L., D.B. Nikolov, and S.K. Burley, *Co-crystal structure of TBP recognizing the minor groove of a TATA element*. Nature, 1993. **365**(6446): p. 520-7.
313. Kim, Y., et al., *Crystal structure of a yeast TBP/TATA-box complex*. Nature, 1993. **365**(6446): p. 512-20.
314. Kokubo, T., et al., *Interaction between the N-terminal domain of the 230-kDa subunit and the TATA box-binding subunit of TFIID negatively regulates TATA-box binding*. Proc Natl Acad Sci U S A, 1994. **91**(9): p. 3520-4.

315. Anandapadamanaban, M., et al., *High-resolution structure of TBP with TAF1 reveals anchoring patterns in transcriptional regulation*. Nat Struct Mol Biol, 2013. **20**(8): p. 1008-14.
316. Kokubo, T., et al., *The yeast TAF145 inhibitory domain and TFIIA competitively bind to TATA-binding protein*. Mol Cell Biol, 1998. **18**(2): p. 1003-12.
317. Gupta, K., et al., *Architecture of TAF11/TAF13/TBP complex suggests novel regulation properties of general transcription factor TFIID*. Elife, 2017. **6**.
318. Grunberg, S. and S. Hahn, *Structural insights into transcription initiation by RNA polymerase II*. Trends Biochem Sci, 2013. **38**(12): p. 603-11.
319. Sainsbury, S., C. Bernecky, and P. Cramer, *Structural basis of transcription initiation by RNA polymerase II*. Nat Rev Mol Cell Biol, 2015. **16**(3): p. 129-43.
320. Imbalzano, A.N., K.S. Zaret, and R.E. Kingston, *Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA*. J Biol Chem, 1994. **269**(11): p. 8280-6.
321. Ranish, J.A. and S. Hahn, *The yeast general transcription factor TFIIA is composed of two polypeptide subunits*. J Biol Chem, 1991. **266**(29): p. 19320-7.
322. DeJong, J. and R.G. Roeder, *A single cDNA, hTFIIA/alpha, encodes both the p35 and p19 subunits of human TFIIA*. Genes Dev, 1993. **7**(11): p. 2220-34.
323. Geiger, J.H., et al., *Crystal structure of the yeast TFIIA/TBP/DNA complex*. Science, 1996. **272**(5263): p. 830-6.
324. Tan, S., et al., *Crystal structure of a yeast TFIIA/TBP/DNA complex*. Nature, 1996. **381**(6578): p. 127-51.
325. Bleichenbacher, M., S. Tan, and T.J. Richmond, *Novel interactions between the components of human and yeast TFIIA/TBP/DNA complexes*. J Mol Biol, 2003. **332**(4): p. 783-93.
326. Liu, D., et al., *Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP*. Cell, 1998. **94**(5): p. 573-83.
327. Ha, I., W.S. Lane, and D. Reinberg, *Cloning of a human gene encoding the general transcription initiation factor IIB*. Nature, 1991. **352**(6337): p. 689-95.

328. Pinto, I., D.E. Ware, and M. Hampsey, *The yeast SUA7 gene encodes a homolog of human transcription factor TFIIB and is required for normal start site selection in vivo*. Cell, 1992. **68**(5): p. 977-88.
329. Deng, W. and S.G. Roberts, *A core promoter element downstream of the TATA box that is recognized by TFIIB*. Genes Dev, 2005. **19**(20): p. 2418-23.
330. Ha, I., et al., *Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II*. Genes Dev, 1993. **7**(6): p. 1021-32.
331. Buratowski, S. and H. Zhou, *Functional domains of transcription factor TFIIB*. Proc Natl Acad Sci U S A, 1993. **90**(12): p. 5633-7.
332. Sainsbury, S., J. Niesser, and P. Cramer, *Structure and function of the initially transcribing RNA polymerase II-TFIIB complex*. Nature, 2013. **493**(7432): p. 437-40.
333. Bushnell, D.A., et al., *Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 Angstroms*. Science, 2004. **303**(5660): p. 983-8.
334. Kostrewa, D., et al., *RNA polymerase II-TFIIB structure and mechanism of transcription initiation*. Nature, 2009. **462**(7271): p. 323-30.
335. Liu, X., et al., *Structure of an RNA polymerase II-TFIIB complex and the transcription initiation mechanism*. Science, 2010. **327**(5962): p. 206-9.
336. Cabart, P., et al., *Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor TFIIB in early elongation complexes*. Proc Natl Acad Sci U S A, 2011. **108**(38): p. 15786-91.
337. Zhang, Z., et al., *Rapid dynamics of general transcription factor TFIIB binding during preinitiation complex assembly revealed by single-molecule analysis*. Genes Dev, 2016. **30**(18): p. 2106-2118.
338. Henry, N.L., M.H. Sayre, and R.D. Kornberg, *Purification and characterization of yeast RNA polymerase II general initiation factor g*. J Biol Chem, 1992. **267**(32): p. 23388-92.
339. Burton, Z.F., et al., *RAP30/74: a general initiation factor that binds to RNA polymerase II*. Mol Cell Biol, 1988. **8**(4): p. 1602-13.

340. Conaway, R.C., et al., *Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex*. Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6205-9.
341. Tan, S., et al., *Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II*. J Biol Chem, 1994. **269**(41): p. 25684-91.
342. Fishburn, J. and S. Hahn, *Architecture of the yeast RNA polymerase II open complex and regulation of activity by TFIIIF*. Mol Cell Biol, 2012. **32**(1): p. 12-25.
343. Khapersky, D.A., et al., *Functions of Saccharomyces cerevisiae TFIIIF during transcription start site utilization*. Mol Cell Biol, 2008. **28**(11): p. 3757-66.
344. Pan, G. and J. Greenblatt, *Initiation of transcription by RNA polymerase II is limited by melting of the promoter DNA in the region immediately upstream of the initiation site*. J Biol Chem, 1994. **269**(48): p. 30101-4.
345. Ohkuma, Y., et al., *Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterization of general transcription factor TFIIIE*. Proc Natl Acad Sci U S A, 1990. **87**(23): p. 9163-7.
346. Inostroza, J., O. Flores, and D. Reinberg, *Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of general transcription factor IIE*. J Biol Chem, 1991. **266**(14): p. 9304-8.
347. Peterson, M.G., et al., *Structure and functional properties of human general transcription factor IIE*. Nature, 1991. **354**(6352): p. 369-73.
348. Okamoto, T., et al., *Analysis of the role of TFIIIE in transcriptional regulation through structure-function studies of the TFIIIEbeta subunit*. J Biol Chem, 1998. **273**(31): p. 19866-76.
349. Maxon, M.E., J.A. Goodrich, and R. Tjian, *Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance*. Genes Dev, 1994. **8**(5): p. 515-24.
350. Ohkuma, Y. and R.G. Roeder, *Regulation of TFIIH ATPase and kinase activities by TFIIIE during active initiation complex formation*. Nature, 1994. **368**(6467): p. 160-3.
351. Gibbons, B.J., et al., *Subunit architecture of general transcription factor TFIIH*. Proc Natl Acad Sci U S A, 2012. **109**(6): p. 1949-54.

352. Schilbach, S., et al., *Structures of transcription pre-initiation complex with TFIID and Mediator*. Nature, 2017. **551**(7679): p. 204-209.
353. Greber, B.J., et al., *The cryo-electron microscopy structure of human transcription factor IID*. Nature, 2017. **549**(7672): p. 414-417.
354. He, Y., et al., *Structural visualization of key steps in human transcription initiation*. Nature, 2013. **495**(7442): p. 481-6.
355. Nair, D., Y. Kim, and L.C. Myers, *Mediator and TFIID govern carboxyl-terminal domain-dependent transcription in yeast extracts*. J Biol Chem, 2005. **280**(40): p. 33739-48.
356. Fishburn, J., et al., *Double-stranded DNA translocase activity of transcription factor TFIID and the mechanism of RNA polymerase II open complex formation*. Proc Natl Acad Sci U S A, 2015. **112**(13): p. 3961-6.
357. He, Y., et al., *Near-atomic resolution visualization of human transcription promoter opening*. Nature, 2016. **533**(7603): p. 359-65.
358. Plaschka, C., et al., *Transcription initiation complex structures elucidate DNA opening*. Nature, 2016. **533**(7603): p. 353-8.
359. Tomko, E.J., et al., *TFIID generates a six-base-pair open complex during RNAP II transcription initiation and start-site scanning*. Nat Struct Mol Biol, 2017. **24**(12): p. 1139-1145.
360. Jeronimo, C. and F. Robert, *The Mediator Complex: At the Nexus of RNA Polymerase II Transcription*. Trends Cell Biol, 2017. **27**(10): p. 765-783.
361. Asturias, F.J., et al., *Conserved structures of mediator and RNA polymerase II holoenzyme*. Science, 1999. **283**(5404): p. 985-7.
362. Davis, J.A., et al., *Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction*. Mol Cell, 2002. **10**(2): p. 409-15.
363. Bernecky, C., et al., *Molecular architecture of the human Mediator-RNA polymerase II-TFIID assembly*. PLoS Biol, 2011. **9**(3): p. e1000603.
364. Plaschka, C., et al., *Architecture of the RNA polymerase II-Mediator core initiation complex*. Nature, 2015. **518**(7539): p. 376-80.

365. Baek, H.J., Y.K. Kang, and R.G. Roeder, *Human Mediator enhances basal transcription by facilitating recruitment of transcription factor IIB during preinitiation complex assembly*. J Biol Chem, 2006. **281**(22): p. 15172-81.
366. Johnson, K.M. and M. Carey, *Assembly of a mediator/TFIID/TFIIA complex bypasses the need for an activator*. Curr Biol, 2003. **13**(9): p. 772-7.
367. Jishage, M., et al., *Transcriptional regulation by Pol II(G) involving mediator and competitive interactions of Gdown1 and TFIIIF with Pol II*. Mol Cell, 2012. **45**(1): p. 51-63.
368. Esnault, C., et al., *Mediator-dependent recruitment of TFIIH modules in preinitiation complex*. Mol Cell, 2008. **31**(3): p. 337-46.
369. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture*. Nature, 2010. **467**(7314): p. 430-5.
370. Grunberg, S., et al., *Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters*. EMBO J, 2016. **35**(22): p. 2435-2446.
371. Allen, B.L. and D.J. Taatjes, *The Mediator complex: a central integrator of transcription*. Nat Rev Mol Cell Biol, 2015. **16**(3): p. 155-66.
372. Adelman, K. and J.T. Lis, *Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans*. Nat Rev Genet, 2012. **13**(10): p. 720-31.
373. Kao, S.Y., et al., *Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product*. Nature, 1987. **330**(6147): p. 489-93.
374. Bentley, D.L. and M. Groudine, *A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL60 cells*. Nature, 1986. **321**(6071): p. 702-6.
375. Gilmour, D.S. and J.T. Lis, *RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in Drosophila melanogaster cells*. Mol Cell Biol, 1986. **6**(11): p. 3984-9.
376. Rougvie, A.E. and J.T. Lis, *The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged*. Cell, 1988. **54**(6): p. 795-804.

377. Plet, A., D. Eick, and J.M. Blanchard, *Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns*. *Oncogene*, 1995. **10**(2): p. 319-28.
378. Krumm, A., et al., *The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region*. *Genes Dev*, 1992. **6**(11): p. 2201-13.
379. Core, L.J., J.J. Waterfall, and J.T. Lis, *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters*. *Science*, 2008. **322**(5909): p. 1845-8.
380. Min, I.M., et al., *Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells*. *Genes Dev*, 2011. **25**(7): p. 742-54.
381. Guo, J. and D.H. Price, *RNA polymerase II transcription elongation control*. *Chem Rev*, 2013. **113**(11): p. 8583-603.
382. Chen, F.X., E.R. Smith, and A. Shilatifard, *Born to run: control of transcription elongation by RNA polymerase II*. *Nat Rev Mol Cell Biol*, 2018. **19**(7): p. 464-478.
383. Kininis, M., et al., *Genomic analyses of transcription factor binding, histone acetylation, and gene expression reveal mechanistically distinct classes of estrogen-regulated promoters*. *Mol Cell Biol*, 2007. **27**(14): p. 5090-104.
384. Rasmussen, E.B. and J.T. Lis, *In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes*. *Proc Natl Acad Sci U S A*, 1993. **90**(17): p. 7923-7.
385. Nechaev, S., et al., *Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila*. *Science*, 2010. **327**(5963): p. 335-8.
386. Titov, D.V., et al., *XPB, a subunit of TFIIH, is a target of the natural product triptolide*. *Nat Chem Biol*, 2011. **7**(3): p. 182-8.
387. Chen, F., X. Gao, and A. Shilatifard, *Stably paused genes revealed through inhibition of transcription initiation by the TFIIH inhibitor triptolide*. *Genes Dev*, 2015. **29**(1): p. 39-47.
388. Henriques, T., et al., *Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals*. *Mol Cell*, 2013. **52**(4): p. 517-28.

389. Shao, W. and J. Zeitlinger, *Paused RNA polymerase II inhibits new transcriptional initiation*. Nat Genet, 2017. **49**(7): p. 1045-1051.
390. Marshall, N.F. and D.H. Price, *Control of formation of two distinct classes of RNA polymerase II elongation complexes*. Mol Cell Biol, 1992. **12**(5): p. 2078-90.
391. Marshall, N.F., et al., *Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase*. J Biol Chem, 1996. **271**(43): p. 27176-83.
392. Cheng, B. and D.H. Price, *Analysis of factor interactions with RNA polymerase II elongation complexes using a new electrophoretic mobility shift assay*. Nucleic Acids Res, 2008. **36**(20): p. e135.
393. Missra, A. and D.S. Gilmour, *Interactions between DSIF (DRB sensitivity inducing factor), NELF (negative elongation factor), and the Drosophila RNA polymerase II transcription elongation complex*. Proc Natl Acad Sci U S A, 2010. **107**(25): p. 11301-6.
394. Keegan, B.R., et al., *The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo*. Development, 2002. **129**(7): p. 1623-32.
395. Amleh, A., et al., *Mouse cofactor of BRCA1 (Cobra1) is required for early embryogenesis*. PLoS One, 2009. **4**(4): p. e5034.
396. Wang, X., et al., *NELF potentiates gene transcription in the Drosophila embryo*. PLoS One, 2010. **5**(7): p. e11498.
397. Eissenberg, J.C., et al., *Cdk9 is an essential kinase in Drosophila that is required for heat shock gene expression, histone methylation and elongation factor recruitment*. Mol Genet Genomics, 2007. **277**(2): p. 101-14.
398. Gilchrist, D.A., et al., *NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly*. Genes Dev, 2008. **22**(14): p. 1921-33.
399. Core, L.J., et al., *Defining the status of RNA polymerase at promoters*. Cell Rep, 2012. **2**(4): p. 1025-35.
400. Lu, X., et al., *Multiple P-TEFbs cooperatively regulate the release of promoter-proximally paused RNA polymerase II*. Nucleic Acids Res, 2016. **44**(14): p. 6853-67.
401. Vos, S.M., et al., *Structure of paused transcription complex Pol II-DSIF-NELF*. Nature, 2018. **560**(7720): p. 601-606.



402. Rahl, P.B., et al., *c-Myc regulates transcriptional pause release*. Cell, 2010. **141**(3): p. 432-45.
403. Schaukowitch, K., et al., *Enhancer RNA facilitates NELF release from immediate early genes*. Mol Cell, 2014. **56**(1): p. 29-42.
404. Bernecky, C., et al., *Structure of transcribing mammalian RNA polymerase II*. Nature, 2016. **529**(7587): p. 551-4.
405. Bernecky, C., J.M. Plitzko, and P. Cramer, *Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp*. Nat Struct Mol Biol, 2017. **24**(10): p. 809-815.
406. Qiu, Y. and D.S. Gilmour, *Identification of Regions in the Spt5 Subunit of DRB Sensitivity-inducing Factor (DSIF) That Are Involved in Promoter-proximal Pausing*. J Biol Chem, 2017. **292**(13): p. 5555-5570.
407. Doamekpor, S.K., et al., *Fission yeast RNA triphosphatase reads an Spt5 CTD code*. RNA, 2015. **21**(1): p. 113-23.
408. Yu, M., et al., *RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II*. Science, 2015. **350**(6266): p. 1383-6.
409. Qiu, H., et al., *The Spt4p subunit of yeast DSIF stimulates association of the Paf1 complex with elongating RNA polymerase II*. Mol Cell Biol, 2006. **26**(8): p. 3135-48.
410. Liu, Y., et al., *Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex*. Mol Cell Biol, 2009. **29**(17): p. 4852-63.
411. Zhou, K., et al., *Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5*. Proc Natl Acad Sci U S A, 2009. **106**(17): p. 6956-61.
412. Chen, F.X., et al., *PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II*. Cell, 2015. **162**(5): p. 1003-15.
413. Chen, F.X., et al., *PAF1 regulation of promoter-proximal pause release via enhancer activation*. Science, 2017. **357**(6357): p. 1294-1298.
414. Hu, X., et al., *A Mediator-responsive form of metazoan RNA polymerase II*. Proc Natl Acad Sci U S A, 2006. **103**(25): p. 9506-11.

415. Mullen Davis, M.A., et al., *Functional interactions of the RNA polymerase II-interacting proteins Gdown1 and TFIIIF*. J Biol Chem, 2014. **289**(16): p. 11143-52.
416. Cheng, B., et al., *Functional association of Gdown1 with RNA polymerase II poised on human genes*. Mol Cell, 2012. **45**(1): p. 38-50.
417. Jishage, M., et al., *Architecture of Pol II(G) and molecular mechanism of transcription regulation by Gdown1*. Nat Struct Mol Biol, 2018. **25**(9): p. 859-867.
418. Price, D.H., A.E. Sluder, and A.L. Greenleaf, *Dynamic interaction between a Drosophila transcription factor and RNA polymerase II*. Mol Cell Biol, 1989. **9**(4): p. 1465-75.
419. Kephart, D.D., et al., *Functional analysis of Drosophila factor 5 (TFIIIF), a general transcription factor*. J Biol Chem, 1994. **269**(18): p. 13536-43.
420. Lei, L., et al., *Functions of the N- and C-terminal domains of human RAP74 in transcriptional initiation, elongation, and recycling of RNA polymerase II*. Mol Cell Biol, 1998. **18**(4): p. 2130-42.
421. DeLaney, E. and D.S. Luse, *Gdown1 Associates Efficiently with RNA Polymerase II after Promoter Clearance and Displaces TFIIIF during Transcript Elongation*. PLoS One, 2016. **11**(10): p. e0163649.
422. Mason, P.B. and K. Struhl, *Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo*. Mol Cell, 2005. **17**(6): p. 831-40.
423. Exinger, F. and F. Lacroute, *6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae*. Curr Genet, 1992. **22**(1): p. 9-11.
424. Venkatesh, S. and J.L. Workman, *Histone exchange, chromatin structure and the regulation of transcription*. Nat Rev Mol Cell Biol, 2015. **16**(3): p. 178-89.
425. Clapier, C.R. and B.R. Cairns, *The biology of chromatin remodeling complexes*. Annu Rev Biochem, 2009. **78**: p. 273-304.
426. Chen, Y., et al., *DSIF, the Paf1 complex, and Tat-SF1 have nonredundant, cooperative roles in RNA polymerase II elongation*. Genes Dev, 2009. **23**(23): p. 2765-77.
427. Schweikhard, V., et al., *Transcription factors TFIIIF and TFIIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms*. Proc Natl Acad Sci U S A, 2014. **111**(18): p. 6642-7.

428. Ishibashi, T., et al., *Transcription factors IIS and IIF enhance transcription efficiency by differentially modifying RNA polymerase pausing dynamics*. Proc Natl Acad Sci U S A, 2014. **111**(9): p. 3419-24.
429. Chen, Z.A., et al., *Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry*. EMBO J, 2010. **29**(4): p. 717-26.
430. Luse, D.S., *Rethinking the role of TFIIF in transcript initiation by RNA polymerase II*. Transcription, 2012. **3**(4): p. 156-9.
431. Cojocar, M., et al., *Genomic location of the human RNA polymerase II general machinery: evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription*. Biochem J, 2008. **409**(1): p. 139-47.
432. Churchman, L.S. and J.S. Weissman, *Nascent transcript sequencing visualizes transcription at nucleotide resolution*. Nature, 2011. **469**(7330): p. 368-73.
433. Gu, W. and D. Reines, *Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II*. J Biol Chem, 1995. **270**(19): p. 11238-44.
434. Galburt, E.A., et al., *Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner*. Nature, 2007. **446**(7137): p. 820-3.
435. Shaevitz, J.W., et al., *Backtracking by single RNA polymerase molecules observed at near-base-pair resolution*. Nature, 2003. **426**(6967): p. 684-7.
436. Kireeva, M.L., et al., *Nature of the nucleosomal barrier to RNA polymerase II*. Mol Cell, 2005. **18**(1): p. 97-108.
437. Donahue, B.A., et al., *Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template*. Proc Natl Acad Sci U S A, 1994. **91**(18): p. 8502-6.
438. Jeon, C. and K. Agarwal, *Fidelity of RNA polymerase II transcription controlled by elongation factor TFIIS*. Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13677-82.
439. Cheung, A.C. and P. Cramer, *Structural basis of RNA polymerase II backtracking, arrest and reactivation*. Nature, 2011. **471**(7337): p. 249-53.
440. Wang, D., et al., *Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution*. Science, 2009. **324**(5931): p. 1203-6.

441. Johnson, T.L. and M.J. Chamberlin, *Complexes of yeast RNA polymerase II and RNA are substrates for TFIIS-induced RNA cleavage*. Cell, 1994. **77**(2): p. 217-24.
442. Sijbrandi, R., U. Fiedler, and H.T. Timmers, *RNA polymerase II complexes in the very early phase of transcription are not susceptible to TFIIS-induced exonucleolytic cleavage*. Nucleic Acids Res, 2002. **30**(11): p. 2290-8.
443. Wade, P.A., et al., *A novel collection of accessory factors associated with yeast RNA polymerase II*. Protein Expr Purif, 1996. **8**(1): p. 85-90.
444. Cao, Q.F., et al., *Characterization of the Human Transcription Elongation Factor Rtf1: Evidence for Nonoverlapping Functions of Rtf1 and the Paf1 Complex*. Mol Cell Biol, 2015. **35**(20): p. 3459-70.
445. Van Oss, S.B., C.E. Cucinotta, and K.M. Arndt, *Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation*. Trends Biochem Sci, 2017. **42**(10): p. 788-798.
446. Rondon, A.G., et al., *Molecular evidence indicating that the yeast PAF complex is required for transcription elongation*. EMBO Rep, 2004. **5**(1): p. 47-53.
447. Pokholok, D.K., N.M. Hannett, and R.A. Young, *Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo*. Mol Cell, 2002. **9**(4): p. 799-809.
448. Van Oss, S.B., et al., *The Histone Modification Domain of Paf1 Complex Subunit Rtf1 Directly Stimulates H2B Ubiquitylation through an Interaction with Rad6*. Mol Cell, 2016. **64**(4): p. 815-825.
449. Formosa, T., *The role of FACT in making and breaking nucleosomes*. Biochim Biophys Acta, 2012. **1819**(3-4): p. 247-55.
450. Brewster, N.K., G.C. Johnston, and R.A. Singer, *A bipartite yeast SSRP1 analog comprised of Pob3 and Nhp6 proteins modulates transcription*. Mol Cell Biol, 2001. **21**(10): p. 3491-502.
451. Formosa, T., et al., *Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN*. EMBO J, 2001. **20**(13): p. 3506-17.
452. Ruone, S., A.R. Rhoades, and T. Formosa, *Multiple Nhp6 molecules are required to recruit Spt16-Pob3 to form yFACT complexes and to reorganize nucleosomes*. J Biol Chem, 2003. **278**(46): p. 45288-95.

453. Orphanides, G., et al., *The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins*. Nature, 1999. **400**(6741): p. 284-8.
454. Kireeva, M.L., et al., *Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription*. Mol Cell, 2002. **9**(3): p. 541-52.
455. Belotserkovskaya, R., et al., *FACT facilitates transcription-dependent nucleosome alteration*. Science, 2003. **301**(5636): p. 1090-3.
456. Kaplan, C.D., L. Laprade, and F. Winston, *Transcription elongation factors repress transcription initiation from cryptic sites*. Science, 2003. **301**(5636): p. 1096-9.
457. Xin, H., et al., *yFACT induces global accessibility of nucleosomal DNA without H2A-H2B displacement*. Mol Cell, 2009. **35**(3): p. 365-76.
458. Jeronimo, C., et al., *The Histone Chaperones FACT and Spt6 Restrict H2A.Z from Intragenic Locations*. Mol Cell, 2015. **58**(6): p. 1113-23.
459. Ivanovska, I., et al., *Control of chromatin structure by spt6: different consequences in coding and regulatory regions*. Mol Cell Biol, 2011. **31**(3): p. 531-41.
460. Yoh, S.M., J.S. Lucas, and K.A. Jones, *The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation*. Genes Dev, 2008. **22**(24): p. 3422-34.
461. Uwimana, N., et al., *Bidirectional terminators in Saccharomyces cerevisiae prevent cryptic transcription from invading neighboring genes*. Nucleic Acids Res, 2017. **45**(11): p. 6417-6426.
462. Ardehali, M.B., et al., *Spt6 enhances the elongation rate of RNA polymerase II in vivo*. EMBO J, 2009. **28**(8): p. 1067-77.
463. Endoh, M., et al., *Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro*. Mol Cell Biol, 2004. **24**(8): p. 3324-36.
464. Orphanides, G. and D. Reinberg, *A unified theory of gene expression*. Cell, 2002. **108**(4): p. 439-51.
465. Shuman, S., *Capping enzyme in eukaryotic mRNA synthesis*. Prog Nucleic Acid Res Mol Biol, 1995. **50**: p. 101-29.
466. Varani, G., *A cap for all occasions*. Structure, 1997. **5**(7): p. 855-8.

467. Lewis, J.D., D. Gorlich, and I.W. Mattaj, *A yeast cap binding protein complex (yCBC) acts at an early step in pre-mRNA splicing*. Nucleic Acids Res, 1996. **24**(17): p. 3332-6.
468. Gingras, A.C., B. Raught, and N. Sonenberg, *eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation*. Annu Rev Biochem, 1999. **68**: p. 913-63.
469. Coppola, J.A., A.S. Field, and D.S. Luse, *Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped*. Proc Natl Acad Sci U S A, 1983. **80**(5): p. 1251-5.
470. Jove, R. and J.L. Manley, *In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites*. J Biol Chem, 1984. **259**(13): p. 8513-21.
471. Wen, Y. and A.J. Shatkin, *Transcription elongation factor hSPT5 stimulates mRNA capping*. Genes Dev, 1999. **13**(14): p. 1774-9.
472. Mandal, S.S., et al., *Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II*. Proc Natl Acad Sci U S A, 2004. **101**(20): p. 7572-7.
473. Pei, Y. and S. Shuman, *Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5*. J Biol Chem, 2002. **277**(22): p. 19639-48.
474. Reed, R., *Mechanisms of fidelity in pre-mRNA splicing*. Curr Opin Cell Biol, 2000. **12**(3): p. 340-5.
475. Moore, M.J. and P.A. Sharp, *Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing*. Nature, 1993. **365**(6444): p. 364-8.
476. Kramer, A., *The structure and function of proteins involved in mammalian pre-mRNA splicing*. Annu Rev Biochem, 1996. **65**: p. 367-409.
477. Galej, W.P., *Structural studies of the spliceosome: past, present and future perspectives*. Biochem Soc Trans, 2018.
478. Di Giammartino, D.C., K. Nishida, and J.L. Manley, *Mechanisms and consequences of alternative polyadenylation*. Mol Cell, 2011. **43**(6): p. 853-66.
479. Shi, Y., *Mechanistic insights into precursor messenger RNA splicing by the spliceosome*. Nat Rev Mol Cell Biol, 2017. **18**(11): p. 655-670.

480. Scheres, S.H. and K. Nagai, *CryoEM structures of spliceosomal complexes reveal the molecular mechanism of pre-mRNA splicing*. *Curr Opin Struct Biol*, 2017. **46**: p. 130-139.
481. Fica, S.M. and K. Nagai, *Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine*. *Nat Struct Mol Biol*, 2017. **24**(10): p. 791-799.
482. Beyer, A.L. and Y.N. Osheim, *Splice site selection, rate of splicing, and alternative splicing on nascent transcripts*. *Genes Dev*, 1988. **2**(6): p. 754-65.
483. McCracken, S., et al., *The C-terminal domain of RNA polymerase II couples mRNA processing to transcription*. *Nature*, 1997. **385**(6614): p. 357-61.
484. Gasch, A., et al., *The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains*. *J Biol Chem*, 2006. **281**(1): p. 356-64.
485. Allen, M., et al., *The structure of an FF domain from human HYP4/FBP11*. *J Mol Biol*, 2002. **323**(3): p. 411-6.
486. Goldstrohm, A.C., et al., *The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1*. *Mol Cell Biol*, 2001. **21**(22): p. 7617-28.
487. Emili, A., et al., *Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD*. *RNA*, 2002. **8**(9): p. 1102-11.
488. David, C.J., et al., *The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex*. *Genes Dev*, 2011. **25**(9): p. 972-83.
489. Mayer, A., et al., *Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution*. *Cell*, 2015. **161**(3): p. 541-554.
490. Nojima, T., et al., *Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing*. *Cell*, 2015. **161**(3): p. 526-540.
491. Alexander, R.D., et al., *Splicing-dependent RNA polymerase pausing in yeast*. *Mol Cell*, 2010. **40**(4): p. 582-93.
492. de la Mata, M., et al., *A slow RNA polymerase II affects alternative splicing in vivo*. *Mol Cell*, 2003. **12**(2): p. 525-32.

493. Fong, N., et al., *Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate*. Genes Dev, 2014. **28**(23): p. 2663-76.
494. Jacquier, A., *The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs*. Nat Rev Genet, 2009. **10**(12): p. 833-44.
495. Porrua, O. and D. Libri, *Transcription termination and the control of the transcriptome: why, where and how to stop*. Nat Rev Mol Cell Biol, 2015. **16**(3): p. 190-202.
496. Creamer, T.J., et al., *Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1*. PLoS Genet, 2011. **7**(10): p. e1002329.
497. Pelechano, V., W. Wei, and L.M. Steinmetz, *Extensive transcriptional heterogeneity revealed by isoform profiling*. Nature, 2013. **497**(7447): p. 127-31.
498. Proudfoot, N.J., *How RNA polymerase II terminates transcription in higher eukaryotes*. Trends Biochem Sci, 1989. **14**(3): p. 105-10.
499. Logan, J., et al., *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, 1987. **84**(23): p. 8306-10.
500. Connelly, S. and J.L. Manley, *A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II*. Genes Dev, 1988. **2**(4): p. 440-52.
501. Richard, P. and J.L. Manley, *Transcription termination by nuclear RNA polymerases*. Genes Dev, 2009. **23**(11): p. 1247-69.
502. Mischo, H.E. and N.J. Proudfoot, *Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast*. Biochim Biophys Acta, 2013. **1829**(1): p. 174-85.
503. Kuehner, J.N., E.L. Pearson, and C. Moore, *Unravelling the means to an end: RNA polymerase II transcription termination*. Nat Rev Mol Cell Biol, 2011. **12**(5): p. 283-94.
504. Birse, C.E., et al., *Coupling termination of transcription to messenger RNA maturation in yeast*. Science, 1998. **280**(5361): p. 298-301.
505. Yonaha, M. and N.J. Proudfoot, *Transcriptional termination and coupled polyadenylation in vitro*. EMBO J, 2000. **19**(14): p. 3770-7.



506. Buratowski, S., *Connections between mRNA 3' end processing and transcription termination*. *Curr Opin Cell Biol*, 2005. **17**(3): p. 257-61.
507. Casanal, A., et al., *Architecture of eukaryotic mRNA 3'-end processing machinery*. *Science*, 2017. **358**(6366): p. 1056-1059.
508. Venkataraman, K., K.M. Brown, and G.M. Gilmartin, *Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition*. *Genes Dev*, 2005. **19**(11): p. 1315-27.
509. Gross, S. and C. Moore, *Five subunits are required for reconstitution of the cleavage and polyadenylation activities of Saccharomyces cerevisiae cleavage factor I*. *Proc Natl Acad Sci U S A*, 2001. **98**(11): p. 6080-5.
510. Pearson, E. and C. Moore, *The evolutionarily conserved Pol II flap loop contributes to proper transcription termination on short yeast genes*. *Cell Rep*, 2014. **9**(3): p. 821-8.
511. Holbein, S., et al., *The P-loop domain of yeast Clp1 mediates interactions between CF IA and CPF factors in pre-mRNA 3' end formation*. *PLoS One*, 2011. **6**(12): p. e29139.
512. Zhelkovsky, A., et al., *The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in Saccharomyces cerevisiae*. *RNA*, 2006. **12**(3): p. 435-45.
513. Lidschreiber, M., et al., *The APT complex is involved in non-coding RNA transcription and is distinct from CPF*. *Nucleic Acids Res*, 2018.
514. Kim, M., et al., *Transitions in RNA polymerase II elongation complexes at the 3' ends of genes*. *EMBO J*, 2004. **23**(2): p. 354-64.
515. Ahn, S.H., M. Kim, and S. Buratowski, *Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing*. *Mol Cell*, 2004. **13**(1): p. 67-76.
516. Baejen, C., et al., *Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes*. *Mol Cell*, 2017. **66**(1): p. 38-49 e6.
517. Whitelaw, E. and N. Proudfoot, *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*, 1986. **5**(11): p. 2915-22.

518. Zhang, Z., J. Fu, and D.S. Gilmour, *CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11*. *Genes Dev*, 2005. **19**(13): p. 1572-80.
519. Steinmetz, E.J., et al., *cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II*. *Mol Cell Biol*, 2006. **26**(7): p. 2688-96.
520. Luo, W., A.W. Johnson, and D.L. Bentley, *The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model*. *Genes Dev*, 2006. **20**(8): p. 954-65.
521. Kim, M., et al., *Distinct pathways for snoRNA and mRNA termination*. *Mol Cell*, 2006. **24**(5): p. 723-734.
522. Dengl, S. and P. Cramer, *Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro*. *J Biol Chem*, 2009. **284**(32): p. 21270-9.
523. Pearson, E.L. and C.L. Moore, *Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor Rat1*. *J Biol Chem*, 2013. **288**(27): p. 19750-9.
524. Schaughency, P., J. Merran, and J.L. Corden, *Genome-wide mapping of yeast RNA polymerase II termination*. *PLoS Genet*, 2014. **10**(10): p. e1004632.
525. Dichtl, B., et al., *A role for SSU72 in balancing RNA polymerase II transcription elongation and termination*. *Mol Cell*, 2002. **10**(5): p. 1139-50.
526. Garas, M., B. Dichtl, and W. Keller, *The role of the putative 3' end processing endonuclease Ysh1p in mRNA and snoRNA synthesis*. *RNA*, 2008. **14**(12): p. 2671-84.
527. Plant, K.E., et al., *Strong polyadenylation and weak pausing combine to cause efficient termination of transcription in the human Ggamma-globin gene*. *Mol Cell Biol*, 2005. **25**(8): p. 3276-85.
528. Schulz, D., et al., *Transcriptome surveillance by selective termination of noncoding RNA synthesis*. *Cell*, 2013. **155**(5): p. 1075-87.
529. Thiebaut, M., et al., *Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance*. *Mol Cell*, 2006. **23**(6): p. 853-64.
530. Arigo, J.T., et al., *Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3*. *Mol Cell*, 2006. **23**(6): p. 841-51.

531. Kiss, T., *Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions*. Cell, 2002. **109**(2): p. 145-8.
532. Gudipati, R.K., et al., *Extensive degradation of RNA precursors by the exosome in wild-type cells*. Mol Cell, 2012. **48**(3): p. 409-21.
533. Neil, H., et al., *Widespread bidirectional promoters are the major source of cryptic transcripts in yeast*. Nature, 2009. **457**(7232): p. 1038-42.
534. Wyers, F., et al., *Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase*. Cell, 2005. **121**(5): p. 725-37.
535. Xu, Z., et al., *Bidirectional promoters generate pervasive transcription in yeast*. Nature, 2009. **457**(7232): p. 1033-7.
536. Marquardt, S., D.Z. Hazelbaker, and S. Buratowski, *Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species*. Transcription, 2011. **2**(3): p. 145-154.
537. van Dijk, E.L., et al., *XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast*. Nature, 2011. **475**(7354): p. 114-7.
538. Steinmetz, E.J., et al., *Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase*. Mol Cell, 2006. **24**(5): p. 735-746.
539. Carroll, K.L., et al., *Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements*. RNA, 2007. **13**(3): p. 361-73.
540. Porrua, O., et al., *In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination*. EMBO J, 2012. **31**(19): p. 3935-48.
541. Nedeá, E., et al., *The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes*. Mol Cell, 2008. **29**(5): p. 577-87.
542. Kim, H.D., J. Choe, and Y.S. Seo, *The sen1(+) gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase*. Biochemistry, 1999. **38**(44): p. 14697-710.
543. Mischo, H.E., et al., *Cell-Cycle Modulation of Transcription Termination Factor Sen1*. Mol Cell, 2018. **70**(2): p. 312-326 e7.

544. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast*. Nature, 2003. **425**(6959): p. 737-41.
545. Kulak, N.A., et al., *Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells*. Nat Methods, 2014. **11**(3): p. 319-24.
546. Porrua, O. and D. Libri, *A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast*. Nat Struct Mol Biol, 2013. **20**(7): p. 884-91.
547. Han, Z., D. Libri, and O. Porrua, *Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination*. Nucleic Acids Res, 2017. **45**(3): p. 1355-1370.
548. Soares, L.M., et al., *Determinants of Histone H3K4 Methylation Patterns*. Mol Cell, 2017. **68**(4): p. 773-785 e6.
549. Roy, K., et al., *Common genomic elements promote transcriptional and DNA replication roadblocks*. Genome Res, 2016. **26**(10): p. 1363-1375.
550. Grzechnik, P., M.R. Gdula, and N.J. Proudfoot, *Pcf11 orchestrates transcription termination pathways in yeast*. Genes Dev, 2015. **29**(8): p. 849-61.
551. Steinmetz, E.J. and D.A. Brow, *Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription*. Mol Cell Biol, 2003. **23**(18): p. 6339-49.
552. Lenstra, T.L., et al., *The role of Ctk1 kinase in termination of small non-coding RNAs*. PLoS One, 2013. **8**(12): p. e80495.
553. Skaar, D.A. and A.L. Greenleaf, *The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p*. Mol Cell, 2002. **10**(6): p. 1429-39.
554. Tudek, A., et al., *Molecular basis for coordinating transcription termination with noncoding RNA degradation*. Mol Cell, 2014. **55**(3): p. 467-81.
555. Chinchilla, K., et al., *Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae*. Eukaryot Cell, 2012. **11**(4): p. 417-29.
556. Finkel, J.S., et al., *Sen1p performs two genetically separable functions in transcription and processing of U5 small nuclear RNA in Saccharomyces cerevisiae*. Genetics, 2010. **184**(1): p. 107-18.

557. Lemay, J.F., et al., *The Nrd1-like protein Seb1 coordinates cotranscriptional 3' end processing and polyadenylation site selection*. Genes Dev, 2016. **30**(13): p. 1558-72.
558. Wittmann, S., et al., *The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA*. Nat Commun, 2017. **8**: p. 14861.
559. Lemay, J.F., et al., *The RNA exosome promotes transcription termination of backtracked RNA polymerase II*. Nat Struct Mol Biol, 2014. **21**(10): p. 919-26.
560. Fox, M.J., et al., *The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway*. PLoS Genet, 2015. **11**(2): p. e1004999.
561. Xiang, K., L. Tong, and J.L. Manley, *Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery*. Mol Cell Biol, 2014. **34**(11): p. 1894-910.
562. Nag, A., K. Narsinh, and H.G. Martinson, *The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase*. Nat Struct Mol Biol, 2007. **14**(7): p. 662-9.
563. Zhao, J., L. Hyman, and C. Moore, *Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis*. Microbiol Mol Biol Rev, 1999. **63**(2): p. 405-45.
564. de Vries, H., et al., *Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors*. EMBO J, 2000. **19**(21): p. 5895-904.
565. Zhang, Z. and D.S. Gilmour, *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, 2006. **21**(1): p. 65-74.
566. Mandel, C.R., et al., *Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease*. Nature, 2006. **444**(7121): p. 953-6.
567. West, S., N. Gromak, and N.J. Proudfoot, *Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites*. Nature, 2004. **432**(7016): p. 522-5.
568. Osheim, Y.N., M.L. Sikes, and A.L. Beyer, *EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3' end cleavage of nascent transcripts*. Chromosoma, 2002. **111**(1): p. 1-12.

569. Osheim, Y.N., N.J. Proudfoot, and A.L. Beyer, *EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage*. *Mol Cell*, 1999. **3**(3): p. 379-87.
570. Gromak, N., S. West, and N.J. Proudfoot, *Pause sites promote transcriptional termination of mammalian RNA polymerase II*. *Mol Cell Biol*, 2006. **26**(10): p. 3986-96.
571. Fong, N., et al., *Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition*. *Mol Cell*, 2015. **60**(2): p. 256-67.
572. Skourti-Stathaki, K., N.J. Proudfoot, and N. Gromak, *Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination*. *Mol Cell*, 2011. **42**(6): p. 794-805.
573. Mischo, H.E., et al., *Yeast Sen1 helicase protects the genome from transcription-associated instability*. *Mol Cell*, 2011. **41**(1): p. 21-32.
574. Moreira, M.C., et al., *Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2*. *Nat Genet*, 2004. **36**(3): p. 225-7.
575. Blair, I.P., et al., *A gene for autosomal dominant juvenile amyotrophic lateral sclerosis (ALS4) localizes to a 500-kb interval on chromosome 9q34*. *Neurogenetics*, 2000. **3**(1): p. 1-6.
576. Suraweera, A., et al., *Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation*. *Hum Mol Genet*, 2009. **18**(18): p. 3384-96.
577. Burghes, A.H. and C.E. Beattie, *Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick?* *Nat Rev Neurosci*, 2009. **10**(8): p. 597-609.
578. Andersen, P.R., et al., *The human cap-binding complex is functionally connected to the nuclear RNA exosome*. *Nat Struct Mol Biol*, 2013. **20**(12): p. 1367-76.
579. Gruber, J.J., et al., *Ars2 promotes proper replication-dependent histone mRNA 3' end formation*. *Mol Cell*, 2012. **45**(1): p. 87-98.
580. Hallais, M., et al., *CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing*. *Nat Struct Mol Biol*, 2013. **20**(12): p. 1358-66.

581. Egloff, S., et al., *Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression*. Science, 2007. **318**(5857): p. 1777-9.
582. Richard, P. and T. Kiss, *Integrating snoRNP assembly with mRNA biogenesis*. EMBO Rep, 2006. **7**(6): p. 590-2.
583. Chanfreau, G., et al., *Processing of a dicistronic small nucleolar RNA precursor by the RNA endonuclease Rnt1*. EMBO J, 1998. **17**(13): p. 3726-37.
584. O'Reilly, D., et al., *Human snRNA genes use polyadenylation factors to promote efficient transcription termination*. Nucleic Acids Res, 2014. **42**(1): p. 264-75.
585. Baillat, D., et al., *Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II*. Cell, 2005. **123**(2): p. 265-76.
586. Ezzeddine, N., et al., *A subset of Drosophila integrator proteins is essential for efficient U7 snRNA and spliceosomal snRNA 3'-end formation*. Mol Cell Biol, 2011. **31**(2): p. 328-41.
587. Yamamoto, J., et al., *DSIF and NELF interact with Integrator to specify the correct post-transcriptional fate of snRNA genes*. Nat Commun, 2014. **5**: p. 4263.
588. Narita, T., et al., *NELF interacts with CBC and participates in 3' end processing of replication-dependent histone mRNAs*. Mol Cell, 2007. **26**(3): p. 349-65.
589. Nino, C.A., et al., *mRNA nuclear export in yeast*. Chem Rev, 2013. **113**(11): p. 8523-45.
590. Meinel, D.M., et al., *Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II*. PLoS Genet, 2013. **9**(11): p. e1003914.
591. Tutucci, E. and F. Stutz, *Keeping mRNPs in check during assembly and nuclear export*. Nat Rev Mol Cell Biol, 2011. **12**(6): p. 377-84.
592. Iglesias, N. and F. Stutz, *Regulation of mRNP dynamics along the export pathway*. FEBS Lett, 2008. **582**(14): p. 1987-96.
593. Segref, A., et al., *Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)<sup>+</sup> RNA and nuclear pores*. EMBO J, 1997. **16**(11): p. 3256-71.
594. Gruter, P., et al., *TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus*. Mol Cell, 1998. **1**(5): p. 649-59.

595. Katahira, J., et al., *The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human*. EMBO J, 1999. **18**(9): p. 2593-609.
596. Santos-Rosa, H., et al., *Nuclear mRNA export requires complex formation between Mex67p and Mtr2p at the nuclear pores*. Mol Cell Biol, 1998. **18**(11): p. 6826-38.
597. Jimeno, S., et al., *The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability*. EMBO J, 2002. **21**(13): p. 3526-35.
598. Strasser, K., et al., *TREX is a conserved complex coupling transcription with messenger RNA export*. Nature, 2002. **417**(6886): p. 304-8.
599. Phatnani, H.P., J.C. Jones, and A.L. Greenleaf, *Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome*. Biochemistry, 2004. **43**(50): p. 15702-19.
600. Johnson, S.A., G. Cubberley, and D.L. Bentley, *Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11*. Mol Cell, 2009. **33**(2): p. 215-26.
601. Johnson, S.A., et al., *The export factor Yra1 modulates mRNA 3' end processing*. Nat Struct Mol Biol, 2011. **18**(10): p. 1164-71.
602. Anderson, J.T., et al., *NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability*. Mol Cell Biol, 1993. **13**(5): p. 2730-41.
603. Tran, E.J., et al., *The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA:protein remodeling events*. Mol Cell, 2007. **28**(5): p. 850-9.
604. Dermody, J.L., et al., *Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation*. PLoS One, 2008. **3**(9): p. e3273.
605. Gilbert, W. and C. Guthrie, *The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA*. Mol Cell, 2004. **13**(2): p. 201-12.
606. Saguez, C., et al., *Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation*. Mol Cell, 2008. **31**(1): p. 91-103.
607. Dieppois, G., N. Iglesias, and F. Stutz, *Cotranscriptional recruitment to the mRNA export receptor Mex67p contributes to nuclear pore anchoring of activated genes*. Mol Cell Biol, 2006. **26**(21): p. 7858-70.



608. Gwizdek, C., et al., *Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription*. Proc Natl Acad Sci U S A, 2006. **103**(44): p. 16376-81.
609. Haruki, H., J. Nishikawa, and U.K. Laemmli, *The anchor-away technique: rapid, conditional establishment of yeast mutant phenotypes*. Mol Cell, 2008. **31**(6): p. 925-32.
610. Bataille, A.R. and F. Robert, *Profiling genome-wide histone modifications and variants by ChIP-chip on tiling microarrays in S. cerevisiae*. Methods Mol Biol, 2009. **543**: p. 267-79.
611. Cassart, C., et al., *Distinct requirement of RNA polymerase II CTD phosphorylations in budding and fission yeast*. Transcription, 2012. **3**(5): p. 231-4.
612. Kim, H.J., et al., *mRNA capping enzyme activity is coupled to an early transcription elongation*. Mol Cell Biol, 2004. **24**(14): p. 6184-93.
613. Lahudkar, S., et al., *A novel role for Cet1p mRNA 5'-triphosphatase in promoter proximal accumulation of RNA polymerase II in Saccharomyces cerevisiae*. Genetics, 2014. **196**(1): p. 161-76.
614. Swaney, D.L., et al., *Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation*. Nat Methods, 2013. **10**(7): p. 676-82.
615. Austenaa, L.M., et al., *Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination*. Mol Cell, 2015. **60**(3): p. 460-74.

