



**HAL**  
open science

# Analyse du rôle de la terminaison de la transcription des ARN non-codants dans la régulation de l'expression des gènes

Noûhou El Moctar Haidara

## ► To cite this version:

Noûhou El Moctar Haidara. Analyse du rôle de la terminaison de la transcription des ARN non-codants dans la régulation de l'expression des gènes. Biologie moléculaire. Université Paris-Saclay, 2021. Français. NNT : 2021UPASL010 . tel-03219719

**HAL Id: tel-03219719**

**<https://theses.hal.science/tel-03219719>**

Submitted on 6 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Analyse du rôle de la terminaison de la transcription des ARN non-codants dans la régulation de l'expression des gènes

Analysis of the role of non-coding transcription termination in the regulation of gene expression

## Thèse de doctorat de l'université Paris-Saclay

École doctorale n°577, Structure et dynamique des systèmes du vivant (SDSV)  
Spécialité de doctorat : Science de la vie et de la santé  
Unité de recherche : Institut Jacques Monod, Université de Paris  
13 rue Hélène Brion, 75013 Paris

Référent : Faculté des sciences d'Orsay

Thèse présentée et soutenue à Paris,  
Le 18 Février 2021, par

**Noûhou EL Moctar HAIDARA**

### Composition du Jury

<b>Mme Julie Soutourina</b> Directeur de recherche, CEA/Université Paris Saclay (I2BC)	Présidente
<b>Mme Françoise Stutz</b> Professeur, Université de Genève	Rapporteuse & Examinatrice
<b>Mme Gwenaëli Badis-Breard</b> Chargée de recherche, HDR, Institut Pasteur	Rapporteuse & Examinatrice
<b>M. Lionel Benard</b> Directeur de recherche CNRS, IBPC	Examineur

### Direction de la thèse

<b>Mme Odil Porrua-Fuerte</b> Chargée de recherche CNRS, IJM, HDR, Université de Paris	Directrice de thèse
--	---------------------

**Titre :** Analyse du rôle de la terminaison de la transcription des ARN non-codants dans la régulation de l'expression des gènes

**Mots clés :** Sen1, ARN non-codant, terminaison de la transcription, expression des gènes

La transcription pervasive est un phénomène universel et constitue une source importante d'ARN non-codants. Elle peut s'interférer avec l'expression normale des gènes. Cependant, la transcription pervasive peut aussi jouer un rôle dans la régulation de l'expression des gènes en favorisant la répression des gènes spécifiques par un mécanisme d'interférence transcriptionnelle. L'efficacité de la terminaison de la transcription non-codante peut fortement influencer sa capacité de réguler l'expression des gènes voisins. Cependant, à ce jour, il n'est pas clair s'il existe des mécanismes qui modulent l'efficacité de la terminaison non-codante en réponse aux signaux environnementaux afin de réguler l'expression des gènes. Durant ma thèse, j'ai abordé cette question en me focalisant sur la régulation de l'hélicase Sen1, un facteur clé de la terminaison de la transcription non-codante chez *S. cerevisiae*.

Dans un premier temps, nous avons identifié une phosphorylation sur une thréonine conservée située dans le domaine catalytique de Sen1 et nous avons montré que cette phosphorylation réduit la capacité de liaison de Sen1 à l'ARN et par conséquent affecte la terminaison de la transcription non-codante. Par la suite, nous avons réalisé des analyses transcriptomiques à haute résolution et montré que cette phosphorylation conduit à la répression de l'expression de *ZAP1* codant pour le régulateur central de la réponse au zinc par interférence transcriptionnelle via un ARN non-codant que nous avons baptisé *ZRN1* (Zap1 Repressor Non-coding gene 1). En plus, de nombreux gènes supplémentaires présentent un profil d'expression imitant les conditions d'excès de zinc où l'expression de *ZAP1* est naturellement réprimée, suggérant que la phosphorylation de Sen1 pourrait participer à la régulation de l'expression des gènes impliqués dans l'homéostasie du zinc.

Dans un second temps, en collaboration avec le laboratoire de F. Posas (IRB, Barcelone, Espagne), nous avons identifié des résidus de Sen1 qui peuvent être phosphorylés *in vitro* par la MAP kinase Hog1, un régulateur clé de la réponse au stress osmotique. Nous avons réalisé des analyses transcriptomiques et montré que certains gènes non-codants cibles de Sen1 sont affectés dans la terminaison de la transcription en condition de stress osmotique. Ces défauts de terminaison sont parfois accompagnés de la répression des gènes en aval ou en antisens et donc, pourraient potentiellement être impliqués dans la régulation de ces gènes en réponse au stress. Nos analyses protéomiques ont indiqué une diminution de l'interaction de Sen1 avec le complexe « Mediator » dans la même condition. Ces résultats nous permettent d'imaginer que les défauts de terminaison en condition de stress osmotique pourraient être provoqués par une diminution de l'efficacité de recrutement de Sen1 par le complexe médiateur. Ces résultats bien que préliminaires, pourraient suggérer l'implication de Sen1 dans la régulation de l'expression des gènes dans la réponse au stress osmotique.

Enfin, nous avons analysé le rôle de l'interaction Sen1 avec la phosphatase Glc7, une protéine impliquée dans la terminaison de la transcription de certains gènes non-codants. Nous avons délété le motif d'interaction de Glc7 dans la protéine Sen1 et montré que cette interaction semble être importante pour la terminaison d'un nombre très restreint d'ARN non-codants. Nous avons ensuite montré que la perte de l'interaction de Sen1 avec Glc7 est associée à une augmentation de son interaction avec Nrd1 et Nab3, deux protéines qui interagissent aussi bien avec Sen1 qu'avec les ARN cibles de Sen1. Ces données suggèrent que Glc7 pourrait participer au relâchement de Sen1 de Nrd1 et Nab3 au niveau de certains

**Title:** Analysis of the role of non-coding transcription termination in the regulation of gene expression

**Keywords:** Sen1, non-coding RNA, transcription termination, gene regulation

Pervasive transcription is a universal phenomenon that leads to the production of a multitude of non-coding RNAs. If left uncontrolled, pervasive transcription can be potentially deleterious for normal gene expression. However, non-coding transcription can also play important regulatory roles, for instance by promoting the repression of specific genes by a mechanism of transcriptional interference. The efficiency of transcription termination can strongly influence the regulatory capacity of non-coding transcription events, yet very little is known about the mechanisms modulating the termination of non-coding transcription in response to environmental cues.

During my PhD I have addressed this question by investigating the mechanisms that regulate the activity of the main actor in termination of non-coding transcription in *S. cerevisiae*, the helicase Sen1. We have identified a phosphorylation at a conserved threonine of the catalytic domain of Sen1 and we have shown that this phosphorylation reduces the efficiency of Sen1-mediated termination by interfering with Sen1 interaction with the RNA. Interestingly, we have found that this phosphorylation impairs termination at an unannotated non-coding gene just upstream of the gene encoding the master regulator of Zn homeostasis, Zap1, and thus, repression of *ZAP1* expression by transcriptional interference. We have named this non-coding gene *ZRV1* for Zap1 Repressor Non-coding gene 1. Furthermore, we have found that many additional genes exhibit an expression pattern mimicking conditions of Zn excess, where *ZAP1* is naturally repressed. Taken together, our results support the idea that Sen1 phosphorylation could be involved in the regulation of the expression of genes involved in zinc homeostasis.

In parallel, in collaboration with the laboratory of F. Posas (IRB, Barcelona, Spain), we have identified several residues that can be phosphorylated *in vitro* by the MPK Hog1, the master regulator of the osmotic stress response. We have performed high-resolution transcriptomic analyses and we have shown that certain Sen1-dependent non-coding genes exhibit impaired transcription termination under osmotic stress. Interestingly, in some cases these termination defects occur concomitantly with the repression of downstream or antisense protein-coding genes, suggesting a possible implication in the regulation of these genes in response to stress. Our proteomic analyses indicated a decrease in the interaction of Sen1 with the mediator complex in the same conditions. These results suggest that the termination defects observed under osmotic stress might be due to a decrease in the recruitment of Sen1 by the mediator complex. These preliminary data suggest a possible role for Sen1 in the regulation of gene expression in the response to osmotic stress.

Finally, we analyzed the role of the interaction of Sen1 with the Glc7 phosphatase, an essential protein that has previously been implicated in transcription termination at a subset of non-coding genes. We have deleted the Glc7 interaction motif in Sen1 and we have shown that this interaction is important for transcription termination at a small subset of non-coding genes. In addition, we have shown that the loss of Sen1 interaction with Glc7 is associated with an increase in the interaction with of Sen1 with Nrd1 and Nab3, two proteins that interact with both Sen1 and Sen1 target RNAs. These data suggest that Glc7 may modestly participate in the release of Sen1 from Nrd1 and Nab3 to the RNA, which might allow more efficient transcription termination by Sen1 in some cases.

*A mes défunt(e)s sœurs, Mouneissa et Zéinabou*

*A mon défunt père, Mohamed EL Moctar*

*A mon défunt cousin, Badi alias grand cousin*

*A ma défunte tante, Aissata*

*A mon frère, Mohamed Lamine*

*A ma mère, Fatoumata*

*A mon épouse, Oumou*

*A tous mes frères et sœurs*

*A ma chère Fadimata*

Merci à Françoise Stutz, Gwenaël Badis-Breard, Lionel Benard et Julie Soutourina d'avoir accepté de faire partie de mon jury de thèse. C'est tout un honneur pour moi de vous avoir au sein de ce comité. Vos contributions m'aideront sans doute à construire ma jeune carrière scientifique.

Merci à Domenico Libri de m'avoir accepté dans son laboratoire. J'ai passé un moment de plaisir dans ton laboratoire. Personnellement, j'ai été marqué par ta sobriété. Je garderai de très bons souvenirs de toi.

Grand merci à Odil Porrua, ma directrice de thèse d'avoir accepté de me prendre sous ta responsabilité. Tu as su m'encourager et me motiver dans les moments les plus difficiles aussi bien sur le plan scientifique que personnel. Merci pour tes soutiens. J'ai passé un moment très agréable sous ta direction. Je retiens de toi une personne aux qualités humaines exceptionnelles. J'ai appris énormément de choses à tes côtés et qui m'aideront dans tous les aspects de la vie. Merci de m'avoir transmis ta patience et ton positivisme. « Nin dah goye » pour dire merci en songhoy.

Merci à tous les membres de l'équipe de Domenico Libri, Juan Juan Xie, Umberto Aiello, Griselda Wetzinger, Tommasso Villa, Yves Clement de m'avoir aidé sur des questions techniques ou dépanné avec des réactifs. Encore merci à Tommasso Villa pour toute l'énergie investie pour la prise en charge de nos commandes de radioactivité malgré la crise sanitaire.

Merci à Benoît Palancade, Lionel Benard et Franck Feuerbarch d'avoir accepté de faire partie de mon comité. Vos questions m'ont vraiment aidé à prendre du recul sur mon projet.

Merci à Michel Werner de m'avoir aidé à rejoindre ma famille pendant la crise sanitaire pendant que j'étais bloqué au Mali.

Merci à tous mes frères et sœurs pour leurs encouragements. Vous avez été là pour moi chaque fois que le besoin s'est manifesté. Merci à mon épouse pour ces soutiens et surtout pour sa patience.

Enfin, merci à mes amis Thierno Sow, Abdoul Karim Dembélé, Moussa Diawara, Seydou Ly, Istifanous Ado, Souleymane Diakité et Insaf Farah pour vos soutiens. Vous avez su être là pour moi.

## Liste des publications

- Modulated termination of non-coding transcription partakes in the regulation of gene expression (**en préparation**).

**Nouhou Haidara** & Odil Porrua

- Intrinsic and extrinsic mechanisms cooperate to ensure efficient termination of RNAPIII transcription (**en préparation**).

Juanjuan Xie, Umberto Aiello, Yves Clement, **Nouhou Haidara**, Mathias Girbig, Jana Schmitzova, Vlad Pena, Christoph Mueller, Domenico Libri and Odil Porrua (**en préparation**)

- Termination of non-coding transcription in yeast relies on both an RNA Pol II CTD interaction domain and a CTD-mimicking region in Sen1, EMBO J 2020.

Zhong Han, Olga Jasnovidova, **Nouhou Haidara**, Agnieszka Tudek1, Karel Kubicek, Domenico Libri, Richard Stefl & Odil Porrua.

# Table des matières

<b>1</b>	<b><i>Introduction générale</i></b>	<b>9</b>
1.1	Initiation de la transcription génique	10
1.1.1	Assemblage du complexe de pré-initiation (PIC) :	10
1.1.2	Architecture de la chromatine dans la région promotrice :	16
1.2	L'élongation de la transcription	17
1.2.1	Description générale	17
1.2.2	Remodelage de la chromatine pendant l'élongation :	18
1.2.3	Le CTD de l'ARNpol II :	20
1.3	Terminaison de la transcription :	23
1.3.1	Terminaison de la transcription des gènes codant des protéines :	23
1.3.2	Terminaison de la transcription non-codante	25
1.4	La transcription cachée ou pervasive :	32
1.4.1	Découverte, origine et contrôle de la transcription pervasive :	32
1.4.2	Fonction de la transcription pervasive :	34
1.5	Rôle de la voie NNS dans la régulation de l'expression des gènes	36
<b>2</b>	<b><i>Objectifs du projet et stratégie expérimentale</i></b>	<b>40</b>
<b>3</b>	<b><i>Résultats et discussions</i></b>	<b>42</b>
3.1	La terminaison modulée de la transcription non-codante participe à la régulation de l'expression des gènes	43
3.2	Analyse globale du rôle des phosphorylations de Sen1 dans la réponse au stress	81
3.3	Analyse fonctionnelle de l'interaction de Sen1 avec la phosphatase Glc7	94
<b>4</b>	<b><i>Bibliographie</i></b>	<b>104</b>



# 1 Introduction générale

## Introduction générale :

Tous les organismes vivants, des plus simples aux plus complexes, ont besoin de réguler l'expression des gènes afin de produire les protéines nécessaires aux différentes fonctions biologiques. Cette régulation peut s'opérer à différentes étapes : la transcription, la maturation de l'ARN, le transport des ARNs, la dégradation de l'ARN et la traduction....

Dans la cellule, la plupart des régulations se réalisent au niveau transcriptionnel et cette étape est réalisée par les machines moléculaires appelées l'ARN polymérase (ARNpol) qui catalysent la synthèse d'une molécule d'ARN à partir d'une matrice d'ADN.

Les procaryotes possèdent une seule ARN polymérase capable de synthétiser différentes classes d'ARN tandis que les eucaryotes disposent de trois types de polymérases nucléaires permettant de générer les différents types d'ARN. Ces trois ARNpol sont : ARNpol I, ARNpol II et l'ARNpol III

L'ARNpol I est constituée de 14 sous-unités. Elle transcrit le précurseur de tous les ARN ribosomiques (ARNr) excepté l'ARNr 5S, qui sont des composants essentiels du ribosome, nécessaires à la croissance et à la prolifération des cellules. La synthèse des ARN ribosomiques représentent plus de 50% de l'activité transcriptionnelle cellulaire (révue : Goodfellow and Zomerdijk, 2013).

L'ARNpol II constituée de 12 sous-unités est responsable de la synthèse d'ARN messagers et de diverses classes d'ARN non-codants.

Durant ma thèse, je me suis intéressé à la régulation de la transcription par l'ARNpol II. Ainsi, dans ce manuscrit, je resterai focalisé sur la transcription réalisée par cette polymérase.

Enfin, l'ARNpol III composée de 17 sous-unités assure la transcription de divers petits ARNs non-codants qui sont essentiels dans de multiples fonctions cellulaires, notamment l'ARNr 5S et les ARNs de transfert, ainsi que d'autres. La transcription par l'ARNpol III est hautement régulée, et son activité élevée est considérée comme une caractéristique des cellules cancéreuses (révue : Cabarcas and Schramm, 2011).

La transcription par l'ARNpol II, comme pour les autres ARN polymérases, se fait en trois grandes étapes : initiation, élongation et la terminaison de la transcription. Chacune de ces étapes se réalise de manière coordonnée afin de réguler *in-fine* l'expression des gènes.

### 1.1 Initiation de la transcription génique

L'initiation de la transcription est un point de contrôle critique de régulation de l'expression des gènes. Elle se réalise selon un mécanisme précis qui débute par l'assemblage du complexe de pré-initiation (PIC) et se termine par le relâchement de la polymérase du site d'initiation de la transcription. Toutes ces étapes sont interconnectées afin de permettre une meilleure coordination de l'expression des gènes.

#### 1.1.1 Assemblage du complexe de pré-initiation (PIC) :

Le PIC est composé de l'ARNpol II ainsi que des facteurs de transcription (TF) TFIIB, TFIIE, TFIIID, TFIIF et TFIIH et TFIIA, qui sont connus collectivement sous le nom de facteurs généraux de la transcription. L'assemblage du PIC au niveau du promoteur provoque l'ouverture de la double hélice d'ADN, stimule l'initiation de la synthèse d'ARN et la dissociation de l'ARNpol II nécessaire au démarrage de la transcription processive (révue : Gupta et al., 2016).

Chacun des facteurs généraux joue un rôle précis dans l'assemblage du PIC et l'initiation de la transcription :

Le facteur TFIIID est constitué d'une protéine de liaison à la boîte TATA appelée TBP (« TATA-binding

proteïn ») et de 13-14 protéines accessoires capables de se lier à la TBP connues sous l'appellation de TAFs (« TBP-associated factors »). L'assemblage du PIC commence par la liaison de la TBP à la boîte TATA. Cette dernière est une séquence consensus d'environ 8 pb située à environ 40 à 120 paires de bases en amont du site d'initiation de la transcription (revue : Smale and Kadonaga, 2003). La boîte TATA est conservée de la levure à l'homme (Cavallini et al., 1988). Cependant, tous les gènes ne disposent pas d'une boîte TATA, certains promoteurs ont des éléments consensus à la place du TATA notamment l'élément initiateur et bien divers autres éléments. Ces éléments seront abordés dans la section suivante de ce manuscrit.

Ensuite intervient TFIIA qui n'est pas nécessaire à la transcription basale, permet de stabiliser le complexe ADN-TBP (Imbalzano et al., 1994)

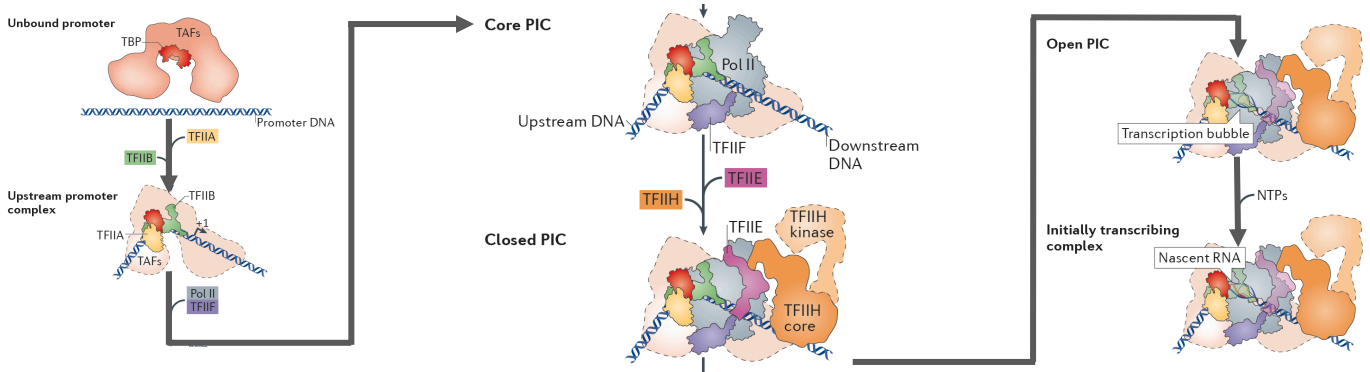
TFIIB s'associe au complexe et permet le recrutement de l'ARNpol II ainsi que la distorsion de l'ADN (Sawadogo and Roeder, 1985; Malik et al., 1991). Elle interagit de manière spécifique avec l'ARNpol II et permet de positionner l'ADN dans le centre catalytique de la polymérase (Chen and Hahn, 2004).

TFIIF s'associe à la polymérase et empêche des interactions non spécifiques de cette dernière avec l'ADN (Conaway et al., 1991), stabilise le PIC (Tan et al., 1994) et permet de maintenir TFIIB au sein du complexe de pré-initiation (Cabart et al., 2011). Par la suite, TFIIF influence la sélection du site d'initiation de la transcription (Ghazy et al., 2004), favorise la formation des liaisons phosphodiester conduisant à la synthèse d'ARN, aide à stabiliser l'hybride ADN/ARN dans le centre catalytique de la ARNpol II (Khapersky et al., 2008). Il a été montré que TFIIF est nécessaire pour l'assemblage du PIC mais ne semble pas être strictement nécessaire à l'initiation de la transcription (Cabart et al., 2011). Enfin, après le départ de l'ARNpol II du promoteur, le complexe d'élongation marque une pause transcriptionnelle, les facteurs TFIIF et TFIIS agissent de concert afin de permettre le redémarrage de l'élongation de la transcription une fois que le blocage est résolu (Schweikhard et al., 2014; revue : Svejstrup, 2007).

Enfin, le recrutement de TFII E et TFII H facilite l'ouverture de l'ADN au niveau du promoteur. Les analyses structurales ont permis de révéler que TFII E est capable d'interagir avec un des brins du duplex d'ADN et conduit ainsi à la transformation du PIC fermé en PIC ouvert. TFII E facilite le recrutement de TFII H au complexe de pré-initiation (Maxon et al., 1994) et sert de pont entre l'ARNpol II et TFII H.

TFII H contient une sous-unité hélicase dépendante de l'hydrolyse de l'ATP, elle transloque le duplex d'ADN devant l'ARNpol II avant de participer à l'ouverture de l'ADN en début de transcription (revue : Hantsche and Cramer, 2017; revue : Sainsbury et al., 2015).

A ce stade, l'ARNpol II est sous forme non phosphorylée. La sous-unité catalytique de TFII H phosphoryle un domaine particulier de l'ARNpol II qui sera décrit postérieurement, ce qui permet le départ de l'ARNpol II du promoteur et l'élongation de la transcription peut ainsi commencer (Wong KH et al., 2014). La figure 1 ci-dessous représente les différentes étapes d'assemblage du PIC.

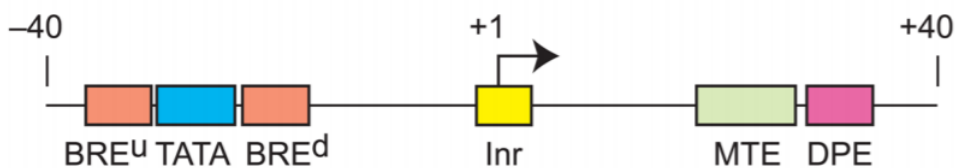


**Figure 1** (adaptée de : Sainsbury et al., 2015) : Assemblage du complexe d'initiation de la transcription *in vitro*

En plus du recrutement de l'ARNpol II, comme mentionné ci-dessus, les facteurs généraux jouent un rôle important dans la reconnaissance du promoteur.

### Le promoteur :

Le promoteur est la séquence minimale requise située en amont d'un gène capable d'initier la transcription. Elle est dotée de séquences particulières ou motifs qui sont reconnus par les facteurs généraux de la transcription permettant ainsi de recruter l'ARNpol II. Au sein de cette séquence, il y'a des motifs de base appelés éléments de base du promoteur qui sont nécessaires pour la transcription *in vitro*. Ces motifs incluent (figure 2) : la boîte TATA, la séquence initiateur (Inr), l'élément de reconnaissance de TFIIB (BRE), le motif de 10 éléments (MTE pour « Motif ten element ») et l'élément en aval du promoteur (DPE pour « downstream promoter element »). Il est important de mentionner que tous les éléments BRE, Inr, MTE et DPE sont absents chez *S. cerevisiae* (revue : Juven-Gershon and Kadonaga, 2010).



**Figure 2** (adaptée de : Juven-Gershon and Kadonaga, 2010) : Séquence du promoteur chez des eucaryotes. Le +1 correspond au site d'initiation de la transcription. La boîte TATA est entourée par des boîtes BRE<sup>u</sup> (« BRE upstream ») et BRE<sup>d</sup> (« BRE downstream »). Les positions -40 et +40 correspondent aux distances par rapport au site d'initiation de la transcription qui est localisé dans la boîte initiateur (Inr).

## La boîte TATA

Elle fut la première à être découverte (Goldberg, 1979). La boîte TATA joue un rôle important dans l'assemblage du PIC, comme mentionné ci-dessus, au niveau du promoteur des gènes (Benoist and Chambon, 1981). Chez *S. cerevisiae*, il est localisé à 40 -120 pb du site d'initiation de la transcription (revue : Smale and Kadonaga, 2003) et cette distance du site d'initiation de la transcription chez l'homme est d'environ 35pb (revue : Smale and Kadonaga, 2003; Yang et al., 2007). Chez *S. cerevisiae*, seuls 20% des gènes possèdent la boîte TATA canonique contre 35% des gènes humains (Gershenzon and Ioshikhes, 2005) et le PIC est recruté via la sous-unité TBP de TFIID (Basehoar et al., 2004). Le motif TATA semble être enrichi dans les gènes qui sont hautement régulés et pour lesquels l'initiation de la transcription fait appel à des remodeleurs de la chromatine tels que les gènes de réponse au stress (Basehoar et al., 2004). Au niveau de ces gènes, la compaction du génome rendant inaccessible le promoteur est diminuée grâce au recrutement des facteurs de transcription capables de modifier localement la structure chromatinienne (Struhl, 1999).

En revanche, il a été révélé que les gènes exprimés de façon constitutive sont déplétées de la boîte TATA canonique ou contiennent un motif dégénéré de la boîte TATA. Pour ces derniers, le recrutement du PIC est aussi dépendant de TFIID. Dans ce cas, la TBP est recrutée sous une forme complexée à TFIID et l'ensemble est positionné par le nucléosome +1 au niveau du promoteur (Rhee and Pugh, 2012).

## L'élément de reconnaissance de TFIIB (BRE)

La boîte BRE a été identifiée comme étant le motif de liaison de TFIIB et est placée en amont de 10 à 30% des gènes qui ont la boîte TATA (Lagrange et al., 1998). En plus, il y'a un second motif BRE de reconnaissance de TFIIB situé en aval de la boîte TATA (Deng, 2005). Ces deux motifs fonctionnent conjointement pour moduler l'expression basale des gènes (Deng, 2005; Lagrange et al., 1998).

## Le motif de 10 éléments (MTE) et le motif en aval du promoteur (DPE) :

Les motifs MTE et DPE situés aux environs de 30pb de la séquence initiatrice, sont généralement retrouvés dans les promoteurs sans boîte TATA. Ils sont fixés par les sous-unités TAF6 et TAF9 de TFIID (revue : Burke and Kadonaga, 1997; Theisen et al., 2010)

## La séquence initiatrice (Inr)

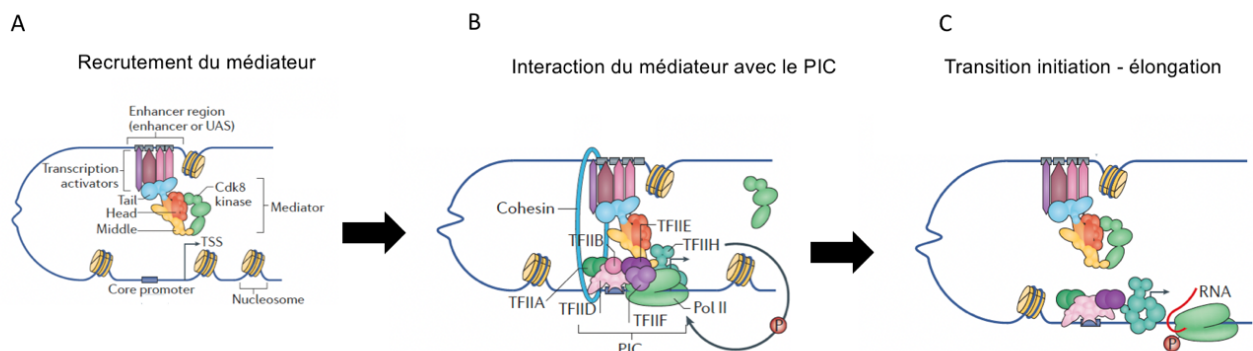
La boîte initiatrice est une séquence consensus incluant le site d'initiation de la transcription. De nombreux facteurs sont capables d'interagir avec cette séquence notamment TFIID via ses sous-unités TafII150et TafII250 (Verrijzer et al., 1995). Chez l'homme, la fréquence de la séquence consensus correspondant à l'initiateur est estimée à 46% (Yang et al., 2007).

Néanmoins, le complexe ARNpol II – facteurs généraux de la transcription requis pour l'expression basale des gènes n'est pas suffisante pour une induction forte de la transcription. Des facteurs spécifiques ou facteurs de transcription sont recrutés afin d'activer ou de réprimer l'expression génique. Pour cela, la cellule fait appel au complexe « Mediator » et au complexe SAGA qui constituent des facteurs clés de l'initiation de la transcription.

## Le complexe « Mediator » :

Le « Mediator » est un complexe multiprotéique conservé chez l'ensemble des eucaryotes. Il sert d'interface physique entre les facteurs de transcription et le PIC. L'assemblage du PIC est nécessaire pour une expression basale des gènes mais ne suffit pas pour induire ou réprimer l'expression des gènes. Pour cela, l'ARNpol II a besoin d'être assistée par des facteurs de transcription spécifiques (activateurs et répresseurs) liés à des séquences spécifiques. Ces facteurs (TF) interagissent avec le complexe « Mediator » afin de transmettre les signaux d'activation ou de répression à la machinerie basale de la transcription (figure 3). Plusieurs études ont rapporté que le « Mediator » est nécessaire pour la transcription de la majorité des gènes codant les protéines (Holstege et al., 1998; Soutourina et al., 2011; Thompson and Young, 1995). Il est recruté au niveau des séquences régulatrices par le biais de son interaction avec les TFs (revue : Poss et al., 2013), ensuite s'ensuit un phénomène de « gene looping » qui permet d'établir des interactions avec les composants du PIC auquel il participe aussi à son assemblage (Cai et al., 2010; Esnault et al., 2008; Eyboulet et al., 2015; Eychenne et al., 2016).

Le « Mediator » est composé de plusieurs sous-unités allant de 25 chez la levure à plus de 30 chez l'homme (revue : Soutourina, 2018) et est capable d'interagir avec une centaine de facteurs de transcription via ses sous-unités pour réguler l'expression des gènes (revue : Poss et al., 2013). Ces différentes sous-unités sont réparties en quatre modules différents : la tête, le milieu, la queue et le module kinase qui n'est pas toujours associé au reste du « Mediator » (Kornberg, 2005). Il a été montré que la tête et le milieu constituent le noyau essentiel du « Mediator » tandis que la queue et le module kinase ont des fonctions dans la régulation. Le « Mediator » participe à la transition initiation-élongation en stimulant la phosphorylation de la serine 5 du CTD de l'ARNpol II et de ce fait, déclenche le départ de l'ARNpol II du promoteur (Kim et al., 1994). Aussi, le « Mediator » est soumis à des régulations de types phosphorylations. En ce sens, il a été montré que la sous-unité Med13 est phosphorylée par la PKA et cette modification permet de moduler l'activité du « Mediator » (Chang et al., 2004). Une autre étude a proposé que la phosphorylation de Med15 participe à la répression des gènes de réponse au stress dans les conditions normales en empêchant le recrutement des activateurs de la transcription au promoteur de ces gènes (Miller et al., 2012).



**Figure 3** (adaptée de : Soutourina, 2018) : recrutement du « Mediator » et association transitoire avec le promoteur.

A : Le « Mediator » ne lie pas directement l'ADN mais interagit avec la chromatine par des intermédiaires : facteurs de transcription liés aux enhancers ou UAS (« Upstream Activator Sequence » chez la levure) et les composants du complexe de pré-initiation (PIC) liés.

B : Ces interactions établies entre différentes régions de la chromatine forment des boucles de chromatine qui amènent les activateurs de transcription à proximité du PIC en formation. Le « Mediator » peut coopérer avec d'autres composants nucléaires telle que la cohésine qui participe à la formation de la boucle.

C : Le départ de l'ARNpol II du promoteur est favorisé par sa phosphorylation sur la serine 5. Cette modification est stimulée par « Mediator ».

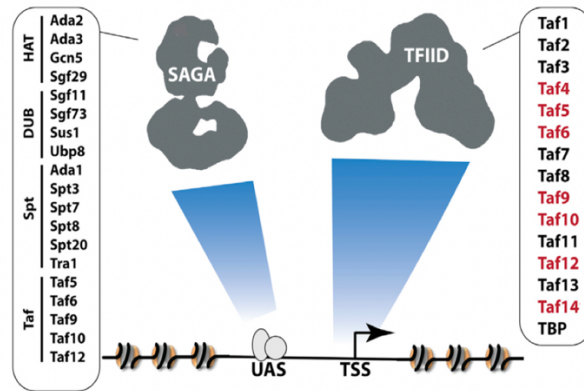
Enfin, Il a été montré que le « Mediator » sert de lien entre la transcription et la réparation de l'ADN. Il participe au recrutement de la machinerie de réparation dans les régions transcrites (revue : André et al., 2020).

### **Le complexe SAGA :**

Dans la région promotrice, l'organisation des nucléosomes constitue une barrière qui empêche l'initiation de la transcription. Ceci indique qu'une déstabilisation importante des interactions ADN/histones est nécessaire pour la formation du PIC et l'activation de la transcription. Pour cela, la cellule recrute des histones acétyltransférases (HAT) pour acétyler les extrémités N-terminales des histones afin de neutraliser leur charge positive, ce qui permet de diminuer l'interaction avec l'ADN chargé négativement. La réversibilité de cette réaction est assurée par l'enzyme HDAC (histones désacétyltransférases) qui catalyse l'enlèvement des groupements acétyles. Le complexe SAGA (Spt-Ada-Gcn5-Acetyltransferase) possède une sous-unité HAT qui l'aide à modifier les histones et favorise ainsi l'activation de la transcription des gènes (revue : Rando and Winston, 2012). Ce complexe a conservé son rôle de co-activateur transcriptionnel tout au long de l'évolution bien que son rôle spécifique dans la transcription ait été modifié chez certaines espèces (Cheon et al., 2020). Il est composé de quatre modules (figure 4) :

1. Le module HAT (histone acétyltransférase) possède une activité histone acétyltransférase portée par la protéine Gcn5. Recrutée au niveau du promoteur des gènes, la protéine Gcn5 assistée par Ada2 et Ada3 acétyle les histones au niveau de la région promotrice afin de permettre à l'ARNpol II d'initier la transcription (Cheon et al., 2020).
1. Le module TAF (« TBP-associated factors ») participe à l'assemblage du PIC en favorisant le recrutement de la TBP. Chez la levure, le module TAF partage la plupart de ces sous-unités avec le complexe TFIID
2. Le module DUB (De-ubiquitination) possède une activité de déubiquitination à travers sa sous-unité Ubp8. Ce module semble plus avoir un rôle dans l'élongation de la transcription. La fonction de ce module sera abordée dans la section suivante.
3. Enfin le module SPT (« Suppressor of Ty insertions ») permet au complexe SAGA d'être recruté par les facteurs de transcription. La sous-unité Tra1 de ce module interagit avec différents activateurs, ce qui conduit au recrutement du complexe SAGA au niveau des promoteurs. Au niveau des gènes où il est recruté, il a été suggéré que ce complexe stimule la formation du complexe de pré-initiation impliquant Spt7 et Spt20 par des interactions avec le TBP et facilite ainsi l'initiation de la transcription (Bonnet et al., 2014; revue : Weake and Workman, 2012).

Des études récentes ont démontré que SAGA et TFIID recrutés au niveau du promoteur des gènes sont responsables de la transcription de la très grande majorité de gènes (Baptista et al., 2017; Warfield et al., 2017) contrairement aux études précédentes (Basehoar et al., 2004; Huisinga and Pugh, 2004). Il n'a pas été formellement démontré que ces deux complexes agissent de concert pour activer la transcription des gènes. Aussi, une interaction physique entre le « Mediator » et le complexe SAGA n'a pas été démontré à ce jour, même si ces deux complexes partagent les mêmes séquences régulatrices (revue : Taatjes, 2017)



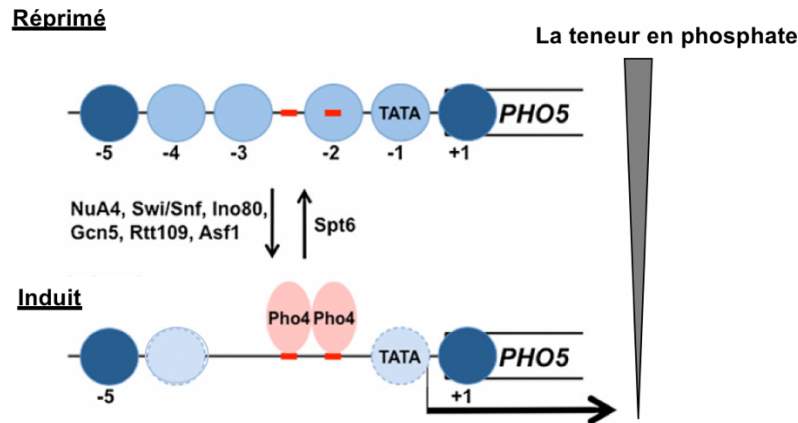
**Figure 4** (adaptée de : Taatjes, 2017) : schéma de la composition et des sites de reconnaissance des complexes SAGA et TFIID.

SAGA reconnaît des séquences spécifiques appelées « upstream activation sequences » (UAS), localisées en amont du site d'initiation de la transcription tandis que TFIID occupe les séquences autour du site d'initiation de la transcription. Les sous-unités de chaque complexe sont indiquées. Les sous-unités du complexe SAGA sont organisées en 4 modules fonctionnels. En rouge, les sous unités du complexe TFIID qui sont présentes en plusieurs copies.

### 1.1.2 Architecture de la chromatine dans la région promotrice :

Dans le noyau, l'ADN génomique est empaqueté pour former la chromatine, qui est composée d'une répétition de sous-unités appelées nucléosomes (Kornberg and Thonmas, 1974; Luger et al., 1997). Chaque nucléosome est composé de 146 pb d'ADN enroulé autour d'un octamère d'histone. Chaque octamère comprend deux exemplaires des histones H2A, H2B, H3 et H4. Une distance d'une cinquantaine de paires de base sépare chaque nucléosome. L'organisation chromatiniennne autour de la région promotrice constitue un obstacle pour le recrutement des facteurs de transcription. Afin de comprendre le mécanisme par lequel la cellule module localement l'organisation de la chromatine autour du promoteur, des études ont été entreprises il y a une vingtaine d'année sur un petit nombre de gènes chez la levure dont le gène *PHO5* pour établir la relation entre la structure de la chromatine et l'initiation de la transcription. Le gène *PHO5* et d'autres gènes régulés par le phosphate sont transcriptionnellement réprimés en cas de fortes teneurs en phosphate et induits en cas de faibles teneurs par les facteurs de transcription Pho2 et Pho4. Les auteurs ont tout d'abord établi le principe selon lequel les nucléosomes sont concentrés dans la région promotrice lorsque les gènes sont réprimés, bloquant l'accès des facteurs de transcription à leurs sites de liaison, et que les nucléosomes sont modifiés ou supprimés lors de l'induction transcriptionnelle (figure 5) (revue : Lohr, 1997). Ensuite d'autres études ont caractérisé en détail la perte de nucléosomes en apportant la preuve qu'elle se produit à des degrés variables au sein d'une population de cellules induites, se propageant à partir du site de liaison de Pho4 (Boeger et al., 2008; Jessen et al., 2006). L'analyse du gène *PHO5* a permis d'établir le concept qu'un grand nombre de facteurs de modification de la chromatine NuA4, Gcn5 et Rtt109 souvent redondants, de complexes de remodelage des nucléosomes (SWI/SNF et INO80) et les chaperons d'histones (Asf1 et Spt6) sont nécessaires pour la modulation de la reconnaissance de la région promotrice (revue : Rando and Winston, 2012).





**Figure 5 :** Régulation du gène *PHO5*.

Représentation du gène *PHO5* lorsqu'il est réprimé (en haut) et induite (en bas). Dans l'état réprimé (forte teneur en phosphate), quatre nucléosomes représentés en bleu clair couvrent la région régulatrice de *PHO5* et bloquent l'un des sites de liaison de Pho4 (ligne rouge). Plusieurs facteurs (listés à gauche), sont requis pour l'induction de l'expression de *PHO5*, qui implique la perte de nucléosomes sur la région de régulation de *PHO5* permettant ainsi le recrutement de Pho4 et l'activation de la transcription. Les lignes pointillées autour des nucléosomes indiquent un degré de perte plus variable. En ajoutant du phosphate, les nucléosomes se réassemblent dans la région de régulation *PHO5* de façon dépendante du Spt6, ce qui entraîne une répression transcriptionnelle.

## 1.2 L'élongation de la transcription

### 1.2.1 Description générale

Après l'ouverture du complexe d'initiation, le premier nucléotide se positionne dans le site catalytique de la polymérase et l'initiation de la transcription est ainsi amorcée. A ce niveau, la synthèse peut être divisée en trois étapes selon la longueur du transcrit qui peut être 5, 10 ou 25 nucléotides (nt). Les transcrits inférieurs à 5 nt sont susceptibles de se dissocier et se libérer de l'ARN polymérase, ce qui conduit à une transcription abortive. Après 5 nt, l'ARNpol II entre dans le site de liaison de TFIIB en provoquant ainsi le déplacement de ce dernier. Lorsque l'ARN naissant atteint une longueur de 10 nt, les contacts étendus entre la polymérase et l'hybride ADN/ARN ainsi formés contribuent à stabiliser le complexe d'élongation (Holstege, 1997). La séparation de la région 5' de l'ARN de la matrice d'ADN permet à la polymérase de quitter le promoteur (Pal and Luse, 2003). Aux alentours de 25 nt, l'initiation de la transcription est finie et l'élongation peut réellement ainsi commencer (Saunders et al., 2006). Les facteurs généraux et les co-activateurs restent associés à la région promotrice et servent d'échafaudage pour une ré-initiation de la transcription. Les facteurs TFIIF et TFIIB sont les seuls à être recrutés pour un nouveau cycle (revue : Hahn, 2004).

La phase d'élongation est une étape d'addition de nucléotides conduisant au rallongement de la molécule d'ARN en cours de synthèse. Cependant, aujourd'hui, il apparaît évident que cette étape est couplée aux différentes étapes de maturation des ARNs (Sims, 2004). La transition initiation – élongation est favorisée par la phosphorylation de la serine 5 du domaine C-terminal (CTD) de l'ARNpol II (Akhtar et al., 2009; Jeronimo and Robert, 2014; Kim et al., 2009; Robinson et al., 2016).

L'élongation de la transcription n'est pas un processus linéaire. Durant la synthèse de la molécule

d'ARN, l'ARNpol II peut rencontrer des obstacles qui peuvent générer une pause ou un ralentissement suivi d'une marche en arrière. Ces pauses peuvent être causées par la séquence de l'ADN, une mauvaise incorporation des nucléotides ou le positionnement des nucléosomes (revue : Wilson et al., 2013) et semblent se concentrer autour des quatre premiers nucléosomes (Churchman and Weissman, 2011). Différents mécanismes sont alors mis en place pour permettre à l'ARNpol II de progresser jusqu'à l'extrémité 3' des gènes afin d'avoir une molécule d'ARN complète. Durant la marche en arrière, l'hybride ADN-ARN formé est maintenu dans le centre catalytique de l'ARNpol II mais l'extrémité 3' de l'ARN sort du site catalytique. Le facteur d'élongation TFIIIS (Dst1) est recruté par la ARNpol II pour réactiver la polymérase en pause. Il induit le clivage de l'ARN afin de replacer l'extrémité 3' dans le centre catalytique de l'ARNpol II, ce qui permet le redémarrage de l'élongation (Cheung and Cramer, 2011; Selth et al., 2010). Cette activité de clivage semble être nécessaire pour la survie cellulaire car le mutant TFIIIS, incapable de cliver n'est pas viable (Sigurdsson et al., 2010). Si la pause persiste, l'ARNpol II peut être polyubiquitinée et dégradée par le protéasome (revue : Svejstrup, 2007).

Au cours de l'élongation, les nucléosomes constituent une barrière physique à la progression de l'ARNpol II. Le remodelage de la chromatine devient ainsi une étape requise pour le déplacement du complexe d'élongation sur la matrice d'ADN. Ce remodelage fait appel aux modifications post-traductionnelles des histones et aux remodeleurs de la chromatine.

### 1.2.2 Remodelage de la chromatine pendant l'élongation :

Durant la transcription, les histones sont sujettes à des modifications de types acétylation, méthylation, ubiquitination et phosphorylation (revue : Kouzarides, 2007). Certaines de ces modifications s'opèrent pendant l'élongation de la transcription afin de maintenir l'accès de l'ARNpol II à la matrice d'ADN durant l'élongation de la transcription (revue : Rando and Winston, 2012). Parmi ces modifications figurent :

#### a) Les acétylations et les ubiquitinations :

Plusieurs études ont montré que les complexes SAGA et NuA4 sont impliqués dans les modifications des histones durant l'élongation de la transcription (Ginsburg et al., 2009; Govind et al., 2007; Wyce et al., 2007). Il a été montré que le double mutant *GCM5* (sous-unité catalytique de SAGA) et *ESA1* (sous-unité de NuA4) présentent un défaut d'élongation de la transcription plus exacerbé suggérant que ces deux complexes agissent de concert pour favoriser l'élongation de la transcription (Ginsburg et al., 2009).

Ainsi, le complexe SAGA en plus de son rôle dans l'initiation de la transcription, est impliqué dans le remodelage de la chromatine durant l'élongation. Il stimule l'acétylation des histones provoquant ainsi l'éviction des histones de la région codante (Govind et al., 2007). En outre, le complexe SAGA possède une activité de déubiquitination qui lui permet de contrôler le niveau d'ubiquitination de l'histone H2B (Wyce et al., 2007). La question sur le rôle de cette ubiquitination durant l'élongation n'a pas été totalement résolue mais néanmoins Il a été proposé que la diminution du niveau d'ubiquitination des histones permet de moduler la dynamique d'assemblage/désassemblage des histones durant l'élongation (Fleming et al., 2008).

Le complexe NuA4 constitué de 13 sous-unités possède aussi une activité acétyltransférase. Il stimule l'acétylation des histones et favorise l'élongation de la transcription de l'ARNpol II à travers le recrutement du complexe RSC (« Chromatine structure remodelling ») un remodeleur de la chromatine qui permet d'assembler et de désassembler les nucléosomes (Carey et al., 2006; Ginsburg et al., 2009; Kasten et al., 2004).

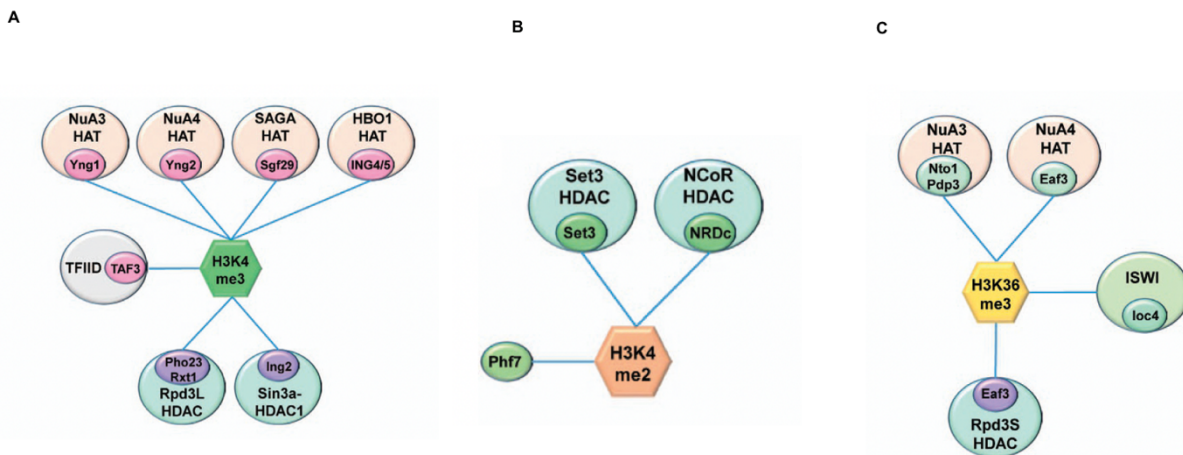
## b) Les méthylations des histones

Contrairement aux acétylations, la méthylation des histones n'affecte pas directement la structure des nucléosomes mais plutôt favorise le recrutement d'autres facteurs. Ces méthylations consistent à ajouter des groupements méthyl aux résidus arginines et lysines des histones (Gary and Clarke, 1998). Ces réactions sont catalysées par les histones méthyl-transférases Set1 et Set2.

L'enzyme Set1 au sein d'un complexe de 7 sous-unités (Bre2, Sdc1, Shg1, Spp1, Swd1, Swd2, and Swd3) appelé le Set1C (Set1 complexe) est responsable de toutes les méthylations de l'histone H3K4 (Briggs, 2001; Miller et al., 2001; Roguev et al., 2001). La méthylation de H3K4 (H3K4me) est une caractéristique des régions 5' des gènes activement transcrits, et cette corrélation suggère ainsi le rôle de Set1 dans les étapes précoces de l'élongation de la transcription (Sayou et al., 2017). Son recrutement par l'ARNpol II phosphorylée sur la serine 5 semble appuyer cette hypothèse (Bae et al., 2020; Ng et al., 2003). Par ailleurs, d'autres études ont révélé le rôle de Set1 dans la répression des gènes par interférence transcriptionnelle (Castelnuovo et al., 2014; Margaritis et al., 2012).

Quant à Set2, il catalyse la méthylation de la lysine 36 de l'histone H3 (H3K36me). Son recrutement est favorisé par la phosphorylation de la serine 2 du CTD de l'ARNpol II (Xiao, 2003). Des expériences d'immunoprécipitation de la chromatine ont permis de montrer que la protéine Set2 et la marque de méthylation de l'histone H3 sont concentrées dans les régions fortement transcrites suggérant le rôle positif de cette méthylation dans la transcription (Krogan et al., 2003). Il arrive aussi que cette modification soit reconnue par le complexe HDAC Rpd3S. Par ce mécanisme, la méthylation co-transcriptionnelle des histones permet la compaction de la chromatine après le passage de la polymérase. La chromatine compactée devient incompétente pour une initiation de la transcription, empêchant ainsi une initiation aberrante dans les régions intragéniques (Millar, 2006; Woo et al., 2017).

Globalement, la méthylation des histones favorise le recrutement des HATs et HDACs permettant ainsi de compacter ou de décompacter la chromatine (figure 6). L'ensemble de ces modifications permet aussi souvent de moduler le recrutement des chaperones d'histones.



**Figure 6** (adaptée de : Woo et al., 2017) : Les effecteurs en aval des méthylations co-transcriptionnelles

A : Une sous-unité distincte (indiquée par des cercles plus petits) des multiples histones acétyltransférases (HAT), y compris les HAT NuA3, NuA4, SAGA et HBO1 et le complexe TFIIID reconnaît la H3K4me3. En outre, les histone désacétylases (HDAC), Rpd3L dans la levure et Sin3a-HDAC1 chez les mammifères interagissent également avec H3K4me3 via des sous-unités spécifiques.

B : Les sous-unités Set3 et NRDC associées aux HDAC se lient préférentiellement à H3K4me2.

C : Les régulateurs de la chromatine qui interagissent avec H3K36me3. Les sous-unités Nto1 et Pdp3 de NuA3 HAT et Eaf3 de NuA4 interagissent avec H3K36me3. En plus, l'HDAC Rpd3S se lie aussi H3K36me3. Enfin, le facteur Loc4, une sous-unité du remodeleur de la chromatine, ISW1 se lie également à H3K36me3.

## Les remodeleurs de la chromatine :

Divers facteurs et complexes enzymatiques ont été mis en évidence dans l'élimination des barrières nucléosomiques au cours de l'élongation de la transcription (revue : Li et al., 2007). Par exemple, il a été montré que les modifications des histones favorisent le recrutement des remodeleurs tels que RSC, SWI/SNF, et le FACT (Dechassa et al., 2010; Hassan et al., 2001; Spain et al., 2014). Alors que le désassemblage des nucléosomes est important pour la transcription, il est également important de restaurer la structure de la chromatine derrière la polymérase pour éviter des initiations aberrantes de la transcription (Kaplan et al., 2003). Les remodeleurs de la chromatine qui sont recrutés de manière co-transcriptionnelle sont impliqués dans l'exercice de cette fonction (revue : Gurard-Levin et al., 2014).

Le complexe RSC (Chromatine structure remodeling) est un complexe de 15 sous-unités importantes pour le déplacement des nucléosomes au cours de l'élongation. Son recrutement dans la région codante des gènes est favorisé par l'acétylation des histones par les complexes SAGA et NuA4 (Carey et al., 2006; Ginsburg et al., 2009) et par la phosphorylation de la serine 2 du CTD de l'ARNpol II (Spain et al., 2014).

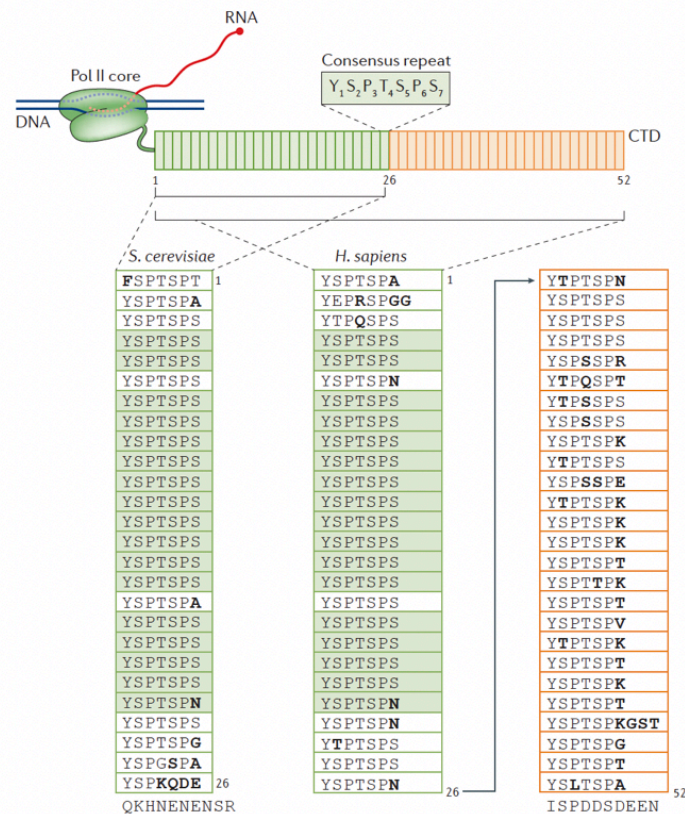
Le complexe SWI/SNF (Switch/Sucrose non-fermentable complex) est composé de 11 sous-unités qui coopèrent pour altérer la structure de la chromatine de manière dépendante de l'hydrolyse de l'ATP (revue : Tang et al., 2010). Il a été proposé que son recrutement aussi est médié par le complexe SAGA (Yoon et al., 2003).

Le complexe FACT (Facilitates chromatine transcription) avec ses deux sous-unités Spt16 et Pob3 est aussi impliqué dans l'altération de la structure de la chromatine au cours de l'élongation de la transcription. Il est recruté par le facteur d'élongation Paf1 afin de retirer les dimères d'histones pour frayer un passage à l'ARNpol II au cours de la transcription (Martin et al., 2018; Squazzo et al., 2002)

### 1.2.3 Le CTD de l'ARNpol II :

L'ARNpol II possède un long domaine protéique non structuré composé de répétitions en tandem d'un heptapeptide avec la séquence consensus YSPTSPS. Ces répétitions forment le domaine carboxy-terminal (CTD) de la plus grande sous-unité de l'ARNpol II (Rpb1). Le nombre de répétitions varie d'une espèce à l'autre allant de 26 chez *S. cerevisiae* à 52 chez l'homme (figure 7)

Le CTD n'est pas nécessaire à l'activité catalytique de l'ARNpol II mais crucial pour les différentes étapes de la transcription : initiation, élongation, terminaison et participe aussi aux processus co-transcriptionnels (Jeronimo et al., 2013) en modulant les interactions avec les différents facteurs au cours de la transcription (Corden, 2013; Hsin and Manley, 2012; Jeronimo et al., 2013). Pour cela, les résidus au sein de la séquence consensus sont soumis à de multiples modifications post-traductionnelles incluant la phosphorylation, l'isomérisation des prolines et la O-glcNAcylation (Corden, 2013; Hsin and Manley, 2012; Jeronimo et al., 2013; Ranuncolo et al., 2012). En plus, les résidus lysines, arginines présentes dans le CTD de certains organismes peuvent être acétylées, méthylées, ubiquitinées (Ranuncolo et al., 2012; Schröder et al., 2013; Sims et al., 2011). L'ensemble de ces modifications constituent le code CTD et jouent un rôle clé dans la transcription mais aussi dans les processus associés à la maturation de l'ARN pré-messagers tels que la mise en place de la coiffe, l'épissage et les modifications de l'extrémité 3' de l'ARN (Heidemann et al., 2013; Hsin and Manley, 2012; Jeronimo et al., 2013).



**Figure 7** (revue : Harlen and Churchman, 2017) : Composition du CTD

Comparaison du CTD de l'homme et de *S. cerevisiae*. Chaque rectangle représente une répétition. Les répétitions qui sont à la fois conservées chez l'homme et la levure sont colorées en vert.

A ce jour, les modifications les mieux étudiées et caractérisées sont les phosphorylations sur la tyrosine 1, la serine 2, la thréonine 4 et les sérines 5 et 7 d'après la séquence (Y1-S2-P3-T4-S5-P6-S7). En général, la phosphorylation du CTD favorise ou empêche la liaison des protéines et permet ainsi de recruter les protéines nécessaires à chaque étape de la transcription. Ces modifications sont portées sur le CTD de manière dynamique et réversible tout au long du cycle transcriptionnel (figure 8A).

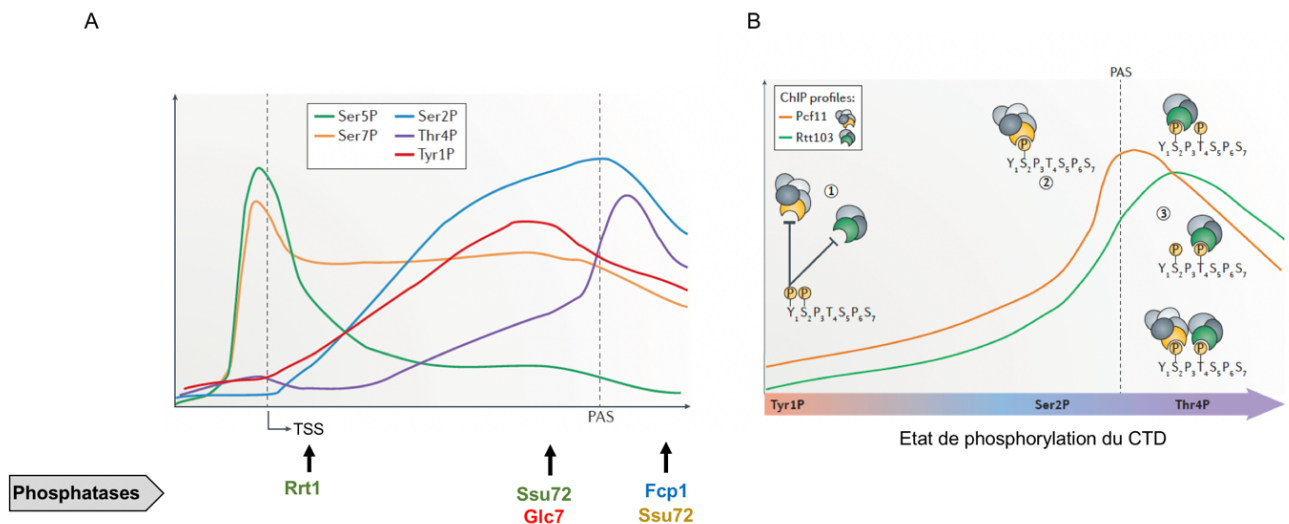
Plusieurs kinases et phosphatases sont responsables de ces modifications. Chez la levure, durant l'initiation de la transcription, la kinase Kin28, une sous-unité du facteur général de transcription TFIIF phosphoryle les sérines 5 (S5P) et 7 (S7P) (Akhtar et al., 2009; Kim et al., 2009). La S5P du CTD aide l'ARNpol II à quitter le promoteur en diminuant l'affinité entre l'ARNpol II et le complexe « Mediator » probablement par un affaiblissement des liaisons d'hydrogènes formées entre les deux complexes (Jeronimo and Robert, 2014; Robinson et al., 2016; Wong et al., 2014). Elle facilite aussi recrutement du complexe enzymatique impliqué dans la mise en place de la coiffe (Komarnitsky, 2000; Rodriguez et al., 2000), une modification requise pour les étapes en aval du processus de maturation de ARNs notamment l'épissage et le clivage en 3' de l'ARN suivi de la polyadénylation (Komarnitsky, 2000; Schroeder et al., 2000). En outre, la S5P permet aussi de recruter la machinerie essentielle pour l'épissage des ARN pré-messagers (Harlen et al., 2016; Morris and Greenleaf, 2000).

Durant, l'élongation, la S5P n'est pas maintenue à un niveau élevé car peu après le départ du promoteur, la phosphatase Rtr1 supprime la marque de phosphorylation de la S5 de l'ARNpol II (Kim et al., 2009; Mosley et al., 2009). La S5P restante est déphosphorylée par la suite par la phosphatase Ssu72, un composant du complexe CPF-CF, qui sera décrit ci-dessous, juste avant la terminaison

(Bataille et al. 2012). Il a été montré que Ssu72 déphosphoryle aussi la sérine 7 (Bataille et al., 2012). La diminution de la S5P est suivie par une augmentation de la S2P due à l'action de la kinase Ctk1 pendant les étapes précoces de l'initiation (Bataille et al., 2012; Qiu et al., 2009). Il a été montré que la S2 est aussi phosphorylée par Bur1 mais cette fois-ci en fin d'élongation (Qiu et al., 2009)

Enfin la phosphatase de la S2P, Fcp1 est nécessaire pour ramener la polymérase à sa forme non-phosphorylée permettant ainsi son recyclage pour un nouveau cycle de transcription (Bataille et al., 2012; Egloff et al. 2012).

Trois modifications semblent agir de concert pour coordonner la terminaison de la transcription (figure 8B). La phosphorylation de la tyrosine 1 par la kinase Slt2 et éventuellement par d'autres kinases non-identifiées (Yurko et al., 2017) empêche le recrutement précoce des facteurs de terminaison (Mayer et al., 2012; Schreieck et al., 2014). En revanche, la déphosphorylation de la tyrosine 1 par la phosphatase Glc7 au niveau du site de polyadénylation favorise le recrutement des facteurs de terminaison (Schreieck et al., 2014). La phosphorylation de la sérine 2 participe aussi au recrutement des sous-unités du complexe CPF-CF (Pcf11 et Rtt103) (Hahn, 2004; Kim et al., 2004; Lunde et al., 2010; Meinhart and Cramer, 2004). Quant à la thréonine 4, une étude récente a montré que la phosphorylation par la kinase Hrr25 contribue au recrutement de Rtt103 pour favoriser la terminaison de la transcription d'une catégorie de snoARNs (Nemec et al., 2019).



**Figure 8** : (adaptée de : Egloff et al., 2012; Harlen and Churchman, 2017) régulation de la transcription par la phosphorylation du CTD

A : Moyenne des profils d'immunoprécipitation de la chromatine (ChIP) des résidus phosphorylés du domaine carboxy-terminal (CTD) de l'ARNpol II à travers les gènes codant des protéines chez *S. cerevisiae*. Le TSS correspond au site d'initiation de la transcription tandis que le PAS correspond au site de polyadénylation.

B : Régulation de la terminaison de la transcription par le CTD. La flèche présente le long de l'axe des x indique l'état de phosphorylation du CTD de l'ARNpol II autour du PAS des gènes codant des protéines. On voit ici le profil de l'immunoprécipitation de la chromatine (ChIP) de deux sous-unités du complexe de terminaison CPF-CF : Pcf11 en orange et Rtt103 en vert. Trois positions le long de la région codante du gène (étapes 1-3) décrivent l'état de phosphorylation du CTD et ses interactions avec le Pcf11 et le Rtt103, suggérant ensemble un modèle pour le recrutement précis des Pcf11 et Rtt103 pendant la transcription. Pendant l'élongation de la transcription (étape 1), la Tyr1 phosphorylée (Tyr1P) empêche le recrutement de Pcf11 et de Rtt103 à l'ARNpol II. La diminution de Tyr1P et l'augmentation de Ser2P (étape 2) favorise le recrutement de Pcf11 à l'ARNpol II autour du PAS. Ser2P et Thr4P recrutent Rtt103 (étape 3). Comme Rtt103 peut se lier à la fois au Ser2P et au Thr4P, les états de Rtt103 liés au CTD sont représentés en aval du PAS. Dans un état, Rtt103 se lie au Ser2P, et dans un autre, le Rtt103 se lie au Thr4P. Dans le troisième état, Pcf11 et Rtt103 peuvent être co-recrutées en se liant aux différentes modifications du CTD pour coordonner le clivage et la polyadénylation de l'ARN.



précisément au niveau d'une région appelée « flap loop ». Cette interaction semble être importante pour une terminaison correcte de la transcription et la maturation de l'ARN naissant (Pearson and Moore, 2014). Afin de terminer la transcription, le recrutement du complexe CPF-CF dans le 3' UTR des gènes est à la fois réalisé par l'interaction avec l'ARNpol II et par la reconnaissance de séquences dans l'ARN naissant. Ensuite, l'endonucléase Ysh1, l'une des sous-unités catalytiques du complexe CPF-CF clive l'ARN au niveau du site de polyadénylation. Le fragment en aval est dégradé par l'exonucléase Rat1 (revue : Kuehner et al., 2011; revue : Porrua and Libri, 2015). Ensuite, la PAP poly(A)-polymérase Pap1 reconnaît le 3' OH du fragment en amont ainsi libéré et catalyse l'addition d'une queue poly(A). L'activité de la protéine Pap1 est régulée par Nab2, une protéine qui interagit avec le complexe CPF-CF. Cette protéine se lie à l'extrémité 3' des ARNs et limite l'addition de résidus adénine à l'extrémité 3' de l'ARN (Hector, 2002). Elle établit un contact avec Mex67 afin de favoriser le transport de l'ARN naissant vers le cytoplasme (Green et al., 2002; Tudek et al., 2018). En plus de son rôle dans la maturation et l'export des ARNs, une étude plus récente vient de montrer le rôle de Nab2 dans l'épissage des ARN pré-messagers (Alpert et al., 2020).

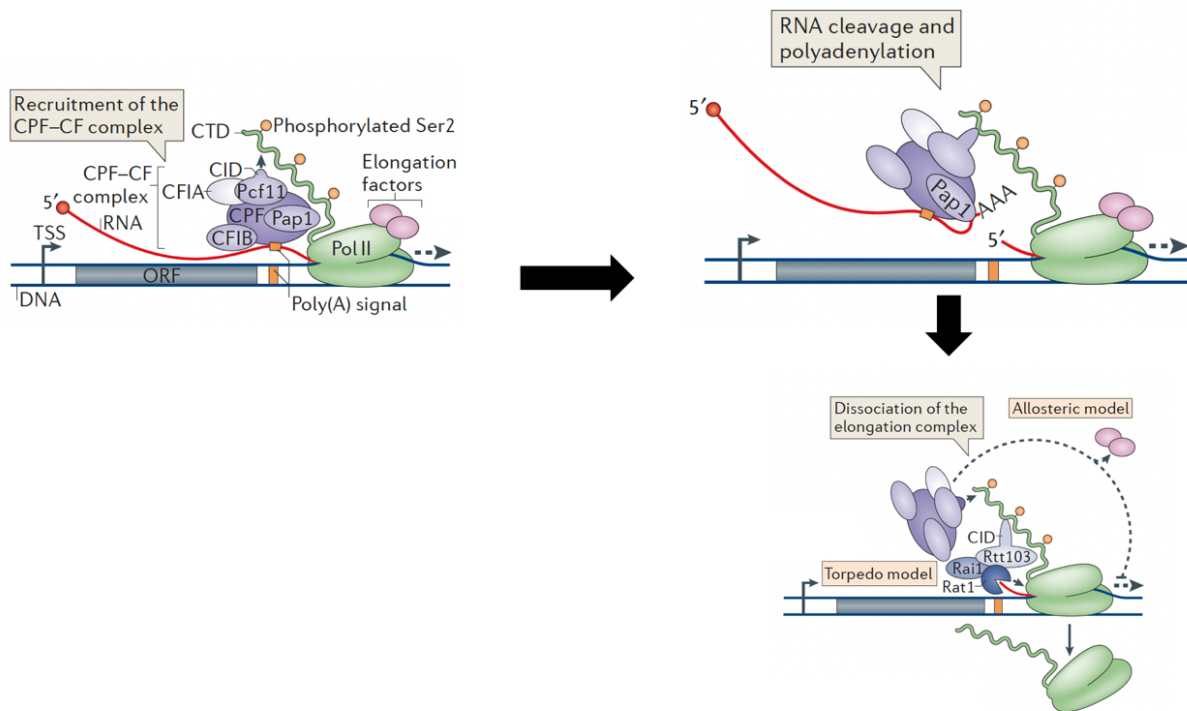
Après le clivage de l'ARN naissant par Ysh1, l'ARNpol II continue à transcrire jusqu'à 150 nucléotides environ en moyenne (Creamer et al., 2011). Le mécanisme de dissociation du complexe d'élongation fait débat au sein de la communauté. En ce sens, deux modèles non-mutuellement exclusifs ont été proposés (figure 10).

Le premier est le modèle allostérique dans lequel le recrutement des facteurs de terminaison au niveau de sites de terminaison par l'ARNpol II provoque un changement conformationnel conduisant à la perte des facteurs d'élongation. Cela s'ensuit par une dissociation de l'ARNpol II de la matrice d'ADN (revue : Kuehner et al., 2011; revue : Porrua and Libri, 2015). Chez l'homme, il a été montré que le site de polyadénylation suffirait pour induire un changement conformationnel du complexe indépendamment du clivage au niveau du 3' de l'ARN (Zhang et al. 2015).

Le deuxième est le modèle torpédo proposant que, après le clivage de l'ARN naissant par Ysh1, suivi par la dégradation du fragment en aval du site de polyadénylation par Rat1 provoque une déstabilisation du complexe d'élongation et la dissociation de l'ARNpol II de la matrice d'ADN (revue : Porrua et al. 2015). En plus, la déplétion de l'endonucléase Ysh1 provoque un défaut de terminaison suggérant ainsi l'importance du clivage dans la terminaison de la transcription (Schaughency et al., 2014). Toujours en faveur du modèle torpédo, au cours de ces dernières années, les analyses génomiques ont permis de valider le rôle de Rat1 à la fois chez l'homme et chez la levure dans la terminaison de la transcription (Baejen et al., 2017; Eaton et al., 2018; Fong et al., 2015) même si la preuve de la terminaison de la transcription par Rat1 *in vitro* fait l'objet de controverse (Dengl and Cramer, 2009; Park et al., 2015; Pearson and Moore, 2013).

Enfin, il n'est pas exclu que ces deux mécanismes peuvent agir de concert afin de terminer efficacement la transcription des gènes codants des protéines (Luo, 2006).





**Figure 10** (adaptée de : Porrua and Libri, 2015a) : Modèle proposé pour la terminaison de la transcription des gènes codant de protéines. Quand la polymérase arrive au niveau du site de polyadénylation (Poly(A)signal), certains composants du complexe CPF-CF reconnaissent des séquences spécifiques dans la région 3' non traduite (UTR) de l'ARN naissant. La sous-unité Pcf11 interagit avec la forme du CTD phosphorylé sur la serine 2 par le biais de son domaine d'interaction avec le CTD (CID) tandis que la sous-unité Ysh1 clive le transcrit au site poly(A). La queue poly(A) est ajoutée par la polymérase Pap1 qui associée au complexe CPF-CF. Enfin, l'extrémité 5' de la partie aval de la transcription est alors dégradée par l'exonucléase Rat1 dans le sens 5'-3'.

### 1.3.2 Terminaison de la transcription non-codante

En plus du complexe CPF-CF, il existe chez *S. cerevisiae* un complexe protéique composé des protéines Nrd1, Nab3 et Sen1 (NNS) qui assure la terminaison d'une fraction importante des ARN non-codants. Comme indiqué précédemment, à la différence de la terminaison par la voie CPF-CF, la terminaison par le complexe NNS est couplée à l'action de l'exosome nucléaire. Le complexe NNS a été initialement décrit pour son rôle dans la terminaison de la transcription de snoARNs (Steinmetz et al., 2001), des composants essentiels de la maturation des ARNs ribosomiques. Plus tard, il a été montré que ce complexe est aussi capable d'induire la terminaison de la transcription de diverses classes d'ARNs non codants. Parmi ces cibles, les ARN non codants appelés CUTs («cryptic unstable transcripts») constituent la fraction majoritaire (Arigo et al., 2006; Thiebaut et al., 2006; Wyers et al., 2005; Schulz et al., 2013) mais une partie de SUTs («stable unannotated transcripts») sont aussi terminés par cette voie (Schulz et al., 2013). L'action de l'exosome nucléaire sur les snoARNs est une étape cruciale pour leur maturation tandis que son action conduit à la dégradation des CUTs.

Le complexe NNS est composé de trois protéines essentielles (revue : Arndt and Reines, 2015):

Nrd1 est une protéine de 63 kDa contenant un motif de reconnaissance de l'ARN (RRM) (Bacikova et al., 2014), un motif d'interaction avec Nab3, l'autre sous-unité du complexe NNS avec laquelle elle forme un hétérodimère et une région N-terminal contenant un domaine de liaison au CTD de l'ARNpol II appelé CID («CTD-Interaction domain») (Conrad et al., 2000; Vasiljeva et al., 2008). Le CID de Nrd1 reconnaît préférentiellement l'isoforme du CTD qui est phosphorylé sur la Ser5 (Kubicek et

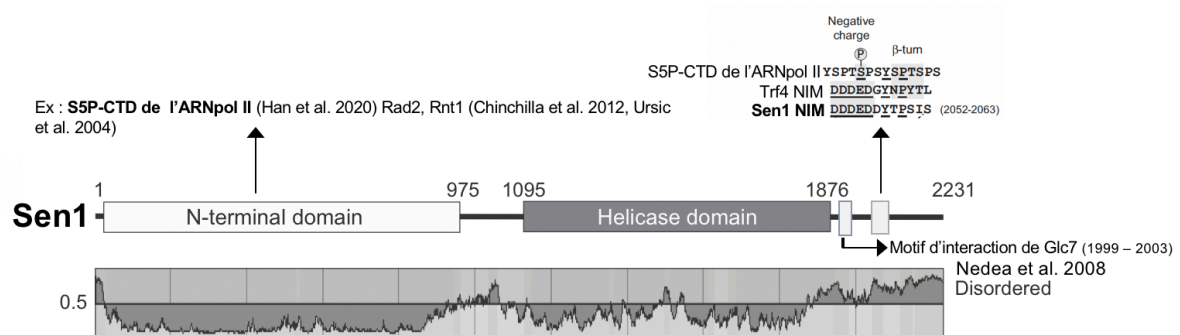
al., 2012; Vasiljeva et al., 2008). La phosphorylation de la serine 5 étant une marque d'élongation précoce de la transcription, l'interaction de Nrd1 avec cette forme phosphorylée du CTD favorise son recrutement précoce et permettrait une terminaison de la transcription plus efficace (Tudek et al., 2014), même si l'interaction de Nrd1 avec le CTD phosphorylé sur la Ser5 n'est pas strictement nécessaire pour la terminaison par la voie NNS (Collin et al., 2019).

La fonction de Nrd1 pourrait être contrôlée par des modifications post-traductionnelles de type phosphorylations et méthylations (Darby et al., 2012; Lee et al., 2020). Il a été proposé que la déphosphorylation de Nrd1 conduit à une relocalisation en réponse à la déplétion du glucose (Darby et al., 2012).

Nab3 est une protéine de 90 kDa qui contient un domaine N-terminal riche en glutamate et en aspartate, un motif central de reconnaissance de l'ARN (RRM) (Wilson et al., 1994) et une région C-terminale essentiellement riche en motif proline et glutamine, une région de faible complexité permettant à la protéine de s'oligomériser. La liaison de la protéine à l'ARN, à travers le motif RRM, pourrait être régulée par la méthylation (Lee et al., 2020). La région C-terminale permet l'oligomérisation de la protéine, ce qui semble être importante pour la terminaison (Carroll et al., 2007; Loya et al., 2013b, 2013a, 2012). Nab3 interagit directement avec Nrd1 et la formation de l'hétérodimère permet une liaison coopérative de ces deux protéines à leurs séquences de reconnaissance respectives (Carroll et al., 2007; Conrad et al., 2000; Hobor et al., 2011).

Nrd1 et Nab3 reconnaissent respectivement des motifs spécifiques GUAA/G et UCUUG dans les ARN cibles (Carroll et al., 2007; Porrua et al., 2012; Schulz et al., 2013; Wlotzka et al., 2011). Ces signaux de terminaison sont souvent associés à des séquences riches en A et U qui permettent une meilleure liaison de ces protéines à leurs séquences cibles et contribuent ainsi à rendre plus efficace la terminaison de la transcription (Porrua et al., 2012).

L'hélicase Sen1 est la seule sous unité du complexe NNS qui possède une activité catalytique. Sen1 est une protéine de 252 kDa, constituée d'un domaine hélicase flanquée par un domaine N-terminal impliqué dans les interactions protéiques (Chinchilla et al., 2012; Han et al., 2020; Ursic et al., 2004) et une région C-terminale non structurée impliquée aussi dans les interactions protéiques (Han et al., 2020; Nedeja et al., 2008; Zhang et al., 2019)(figure 11).



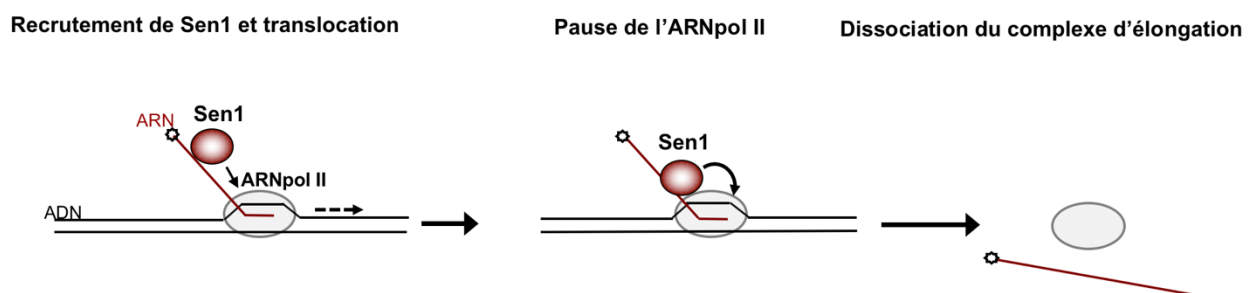
**Figure 11** : Schéma de la séquence de Sen1.

Les domaines globulaires sont indiqués par des barres pleines tandis que les régions intrinsèquement désordonnées sont indiquées par une ligne. En bas du schéma de la protéine, la prédiction de la région désordonnée obtenue à l'aide d'IUPred (Dosztanyi et al., 2005). S5P-CTD de la PolIII, Rad2 et Rnt1 interagissent avec le domaine N-terminal (en haut à gauche). L'interaction identifiée précédemment au laboratoire est indiquée en gras (Han et al., 2020). La séquence interagissant avec Nrd1 (« Nrd1 interacting motif » ou NIM) de Sen1 et Trf4 est détaillée (en haut à droite) : les éléments structuraux importants pour l'interaction avec le CID sont colorés et soulignés. Les acides aminés conservés sont soulignés.

Les hélicases sont des protéines définies à l'origine comme des enzymes qui utilisent la liaison et/ou l'hydrolyse de l'ATP pour favoriser la séparation des deux brins d'une molécule d'acide nucléique (Fairman-Williams et al., 2010; Singleton et al., 2007). Selon leur séquence et leur structure, les hélicases peuvent être classées en six superfamilles (SF). Sen1 appartient à la superfamille 1.

L'hélicase Sen1 est également la seule sous unité du complexe NNS à être conservée dans la plupart des eucaryotes. Son homologue humain, la senataxine, joue également un rôle dans la terminaison de la transcription dans plusieurs modèles cellulaires (Miller et al., 2015; Skourti-Stathaki et al., 2011; Suraweera et al., 2009). Des mutations qui se trouvent plus souvent dans les régions les plus conservées de la senataxine, plus particulièrement le domaine N-terminal et le domaine hélicase, sont liées à la sclérose latérale amyotrophique de type 4 (ALS4) et l'ataxie-apraxie oculaire de type 2 (AOA2). En outre, l'introduction de mutations associées à l'AOA2 dans les résidus équivalents de Sen1 provoque des défauts de terminaison de la transcription à la fois *in vivo* (Chen et al., 2014) et *in vitro* (Leonaitė et al., 2017). Ce qui permet de supposer que ces troubles neurologiques pourraient être dus à des défauts de terminaison.

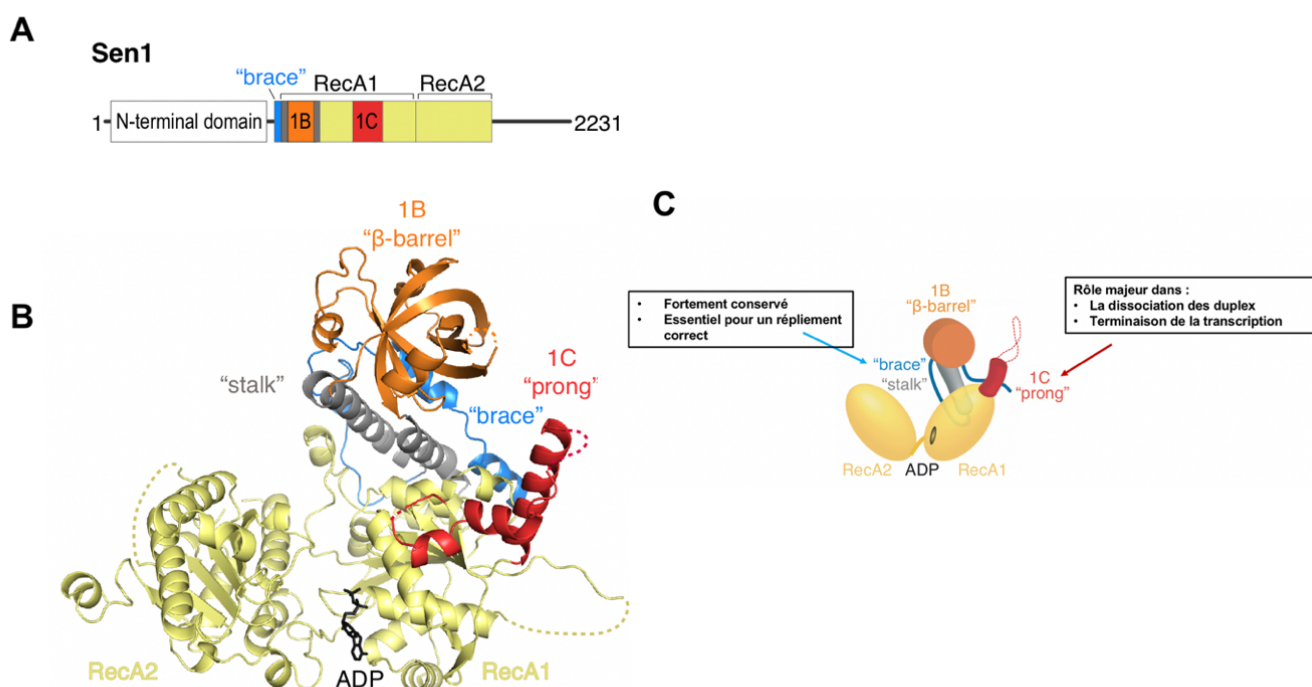
Différentes études réalisées au sein de notre laboratoire ont permis de caractériser en détail le mécanisme de terminaison de la transcription par le complexe NNS. Tout d'abord, il a été montré que l'hélicase Sen1 est responsable de la dissociation du complexe d'élongation (Porrúa and Libri, 2013). Le rôle précis de Nrd1 et Nab3 dans la terminaison reste mal élucidé mais le modèle prédominant postule que ces protéines induisent la pause de l'ARNpol II, ce qui est un prérequis pour la terminaison par la voie dépendante de NNS ainsi que pour les autres voies de terminaison (revue : Porrúa et al., 2016). Ensuite il a été déterminé que le domaine hélicase de Sen1 est suffisant pour la terminaison de la transcription *in vitro* (Han et al., 2017; Leonaitė et al., 2017). Quant au mécanisme employé par Sen1, il implique une translocation de Sen1 sur l'ARN naissant dans le sens 5' - 3' jusqu'à l'ARNpol II. Une fois Sen1 entre en contact avec l'ARNpol II, elle exerce une action mécanique qui, accompagnée par des interactions entre des régions spécifiques de Sen1 et l'ARNpol II, provoque un changement conformationnel au niveau du complexe d'élongation conduisant ainsi à sa dissociation (figure 12)(Han et al., 2017). L'existence de ce changement conformationnel est appuyée par les résultats d'une étude de la réaction de terminaison de la transcription par Sen1 à l'échelle de la molécule unique réalisée au sein de notre laboratoire en collaboration avec l'équipe de Terence Strick dans notre institut (Wang et al., 2019).



**Figure 12 :** Modèle simplifié de terminaison de la transcription par l'hélicase Sen1

Une étude structure – fonction (figure 13) réalisée dans le cadre d'une collaboration entre notre laboratoire et celui d'Elena Conti (Max Planck, Munich) a révélé plusieurs caractéristiques structurales essentielles pour la terminaison de la transcription (Leonaitė et al., 2017). Comme les autres hélicases de la même famille, le domaine hélicase de Sen1 est composé de deux domaines

RecA localisés en tandem qui forment le core (figure 13A). Du premier domaine RecA émergent plusieurs domaines auxiliaires : le « stalk », le « prong », et le « barrel » (figure 13B). En général, le core est conservé au sein des hélicases proches mais les domaines accessoires présentent des caractéristiques distinctes qui confèrent à chaque hélicase des propriétés spécifiques. En effet, les domaines accessoires le « prong » et le « barrel » de Sen1 présentent des caractéristiques structurales différentes par rapport aux domaines équivalents dans Upf1, une hélicase de la même famille qui n'est pas capable d'induire la terminaison de la transcription *in vitro* (Porrúa and Libri, 2013). En plus, il y'a un autre sous-domaine appelé le « brace » qui est spécifique à Sen1 et conservé au cours de l'évolution. Ce sous-domaine est impliqué dans un grand nombre d'interactions intramoléculaires avec d'autres sous-domaines notamment le RecA1, le « stalk » et le « barrel » (figure 13B, C). Ces interactions semblent être nécessaires pour un repliement correct de la protéine. La même étude a montré que le sous-domaine nommé « prong » est essentiel pour la viabilité cellulaire ainsi que pour la terminaison de la transcription *in vivo* et *in vitro* (Leonaitė et al., 2017).



**Figure 13** (adaptée de : Leonaitė et al., 2017) : Structure du domaine hélicase de la protéine Sen1

A : Schéma de la protéine Sen1 avec les régions structurées et non structurées, d'après les prédictions, représentées respectivement par des rectangles et des lignes. Les fragments cristallisés sont mis en évidence en couleur. Les domaines RecA1 et RecA2 sont en jaune, le « stalk » en gris, le « barrel » en orange et le « prong » en rouge. B : La structure du domaine hélicase déterminée en absence d'ARN. Les lignes pointillées indiquent des boucles désordonnées non modélisées dans la structure. C : Schéma simplifié des principales caractéristiques structurales.

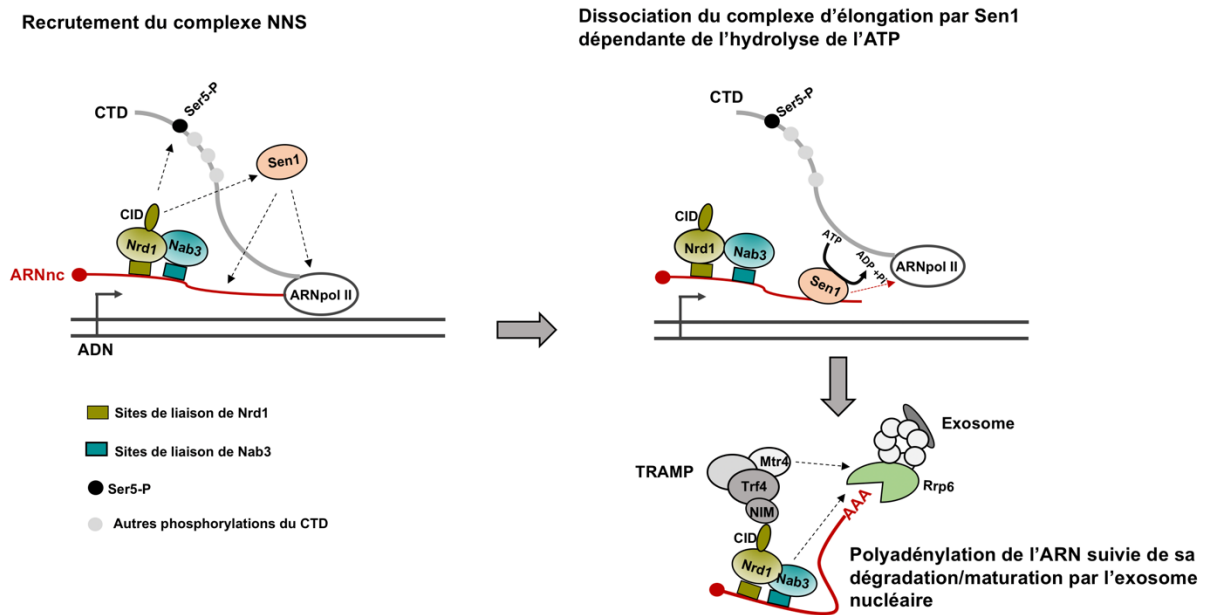
Des approches *in vivo* et *in vitro* ont montré l'incapacité de Sen1 à se lier aux ARNs de façon spécifiques (Creamer et al., 2011; Porrúa and Libri, 2013). Ceci pourrait impliquer que le recrutement de Sen1 doit être régulé afin de spécifier son action sur les ARN non-codants, surtout compte tenu des niveaux relativement faibles de la protéine Sen1 (63-498 molécules/cellule selon les études (Chong et al., 2015; Ghaemmaghami et al., 2003; Newman et al., 2006)). Pendant plusieurs années, il a été proposé que Nrd1 et Nab3 fonctionnent comme des adaptateurs qui apportent la spécificité nécessaire à Sen1. Cependant, jusqu'à très récemment, les régions impliquées dans les interactions entre Sen1 et Nrd1-Nab3 n'étaient pas connues et la question de savoir si ces interactions sont

réellement nécessaires pour le recrutement de Sen1 restait également ouverte. Durant ma thèse, j'ai contribué à une étude qui a abordé ces questions. Ce travail dans lequel je figure comme 3<sup>ème</sup> auteur a été publié par le journal EMBO.

Dans cette étude, nous avons tout d'abord montré que le CID de Nrd1 est nécessaire pour l'interaction de Nrd1 avec Sen1. La même observation avait été rapportée dans une étude indépendante (Heo et al., 2013). Nous avons ensuite identifié un motif dans la région C-terminale de Sen1 qui est très similaire à un motif nommé NIM (« Nrd1 Interaction motif ») qui avait été trouvé précédemment dans la protéine Trf4, un facteur impliqué dans la dégradation des ARN non-codants (Tudek et al., 2014). Ce motif mime le CTD de l'ARNpol II et par conséquent est reconnu par les mêmes surfaces du CID de Nrd1 qui interagissent avec le CTD. Nous avons démontré que le NIM dans la protéine Sen1 est le déterminant principal dans l'interaction de Sen1 avec l'hétérodimère Nrd1 et Nab3 et nous avons fourni les détails structuraux de cette interaction. En revanche, nous avons été surpris de constater que ce motif n'est pas strictement nécessaire à la terminaison de la transcription non-codante bien qu'elle participe au recrutement efficace de Sen1 aux ARNs cibles. En effet, nous avons montré avec des analyses à l'échelle du génome que Sen1 interagit avec la plupart des ARNs cellulaires et que l'interaction de Sen1 avec Nrd1 et Nab3 contribue uniquement à l'association de Sen1 avec les ARN non-codants. Ces données fragilisent le modèle selon lequel Sen1 serait recrutée par Nrd1 et Nab3 afin de terminer efficacement la transcription. Par ailleurs, nous avons montré que le domaine N-terminal de Sen1 interagit avec le CTD de l'ARNpol II phosphorylée sur la Ser5 et cette interaction est importante pour la terminaison de la transcription non-codante. Toutefois, la diminution de la vitesse d'élongation de la polymérase rend le domaine N-terminal de Sen1 moins nécessaire à la terminaison de la transcription. Ce résultat nous a permis de proposer que l'interaction de Sen1 avec le CTD de l'ARNpol II favorise l'action de Sen1 dans un contexte de compétition cinétique entre l'élongation et la terminaison. Enfin, nous avons également montré dans un système hétérologue que le domaine N-terminal de Sen1 interagit avec la région C-terminale, ce qui pourrait moduler l'interaction de Sen1 avec le CTD de l'ARNpol II et Nrd1. Nos conclusions nous permettent de proposer un modèle moléculaire complexe (figure 14) dans lequel différentes interactions protéines-protéines peuvent contrôler la fonction spécifique de Sen1 dans la terminaison de la transcription non-codante (Han et al., 2020).

En plus du NIM, il a été identifié un autre motif dans la région C-terminale de Sen1 qui est impliqué dans l'interaction de Sen1 avec la phosphatase Glc7 (figure 11). Une étude précédente a montré que Glc7 est responsable de la déphosphorylation de la Tyr1 du CTD de l'ARNpol II et que la déplétion de Glc7 provoque des défauts de terminaison de la transcription dans une fraction des gènes non-codants dépendants de la voie NNS (Collin et al., 2019; Schreieck et al., 2014). Durant ma thèse, j'ai tenté de caractériser le rôle de cette interaction entre Sen1 et Glc7. Les résultats préliminaires concernant cette étude sont présentés dans la section des résultats de ce manuscrit.

Une fois que le complexe d'élongation est dissocié par Sen1. Les protéines Nrd1 et Nab3 resteraient associés à l'ARN non-codant. Cela permettrait à ces protéines de recruter le complexe TRAMP et l'exosome afin de permettre la dégradation des ARN nucléaires (Fasken et al., 2015; LaCava et al., 2005; Tudek et al., 2014; Vanáčová et al., 2005; Wyers et al., 2005). La dégradation des ARNs cibles permettrait le relâchement et recyclage des protéines Nrd1 et Nab3 pour un nouveau cycle de terminaison de la transcription (Villa et al., 2020)



**Figure 14 :** Modèle de terminaison de la transcription par le complexe NNS

Au cours de la transcription, Nrd1 serait recrutée par l'ARNpol II via une interaction entre le CID de Nrd1 et la Ser5-P du CTD. Le recrutement de Nrd1 et Nab3 est facilité par la reconnaissance des motifs spécifiques dans l'ARN non-codant. Sen1 semble être recrutée via une liaison aspécifique avec l'ARN non-codant et/ou une interaction avec l'ARNpol II. Le recrutement de Sen1 est aussi favorisé par l'interaction de Sen1 avec le CID de Nrd1. Sen1 transloque par l'ARN naissant et dissocie le complexe d'élongation de manière dépendante de l'hydrolyse de l'ATP. Ensuite, Nrd1 recrute le complexe TRAMP qui polyadényle l'ARN via une interaction entre le CID de Nrd1 et NIM de Trf4. Enfin, Mtr4, une sous-unité du complexe du TRAMP et Nab3 établissent une interaction avec l'exosome par le biais de Rrp6. Ce recrutement de l'exosome favorise la dégradation de l'ARN non-codant dans le cas des CUTs et la maturation dans le cas des snoARNs. Les flèches discontinues indiquent les interactions possibles.

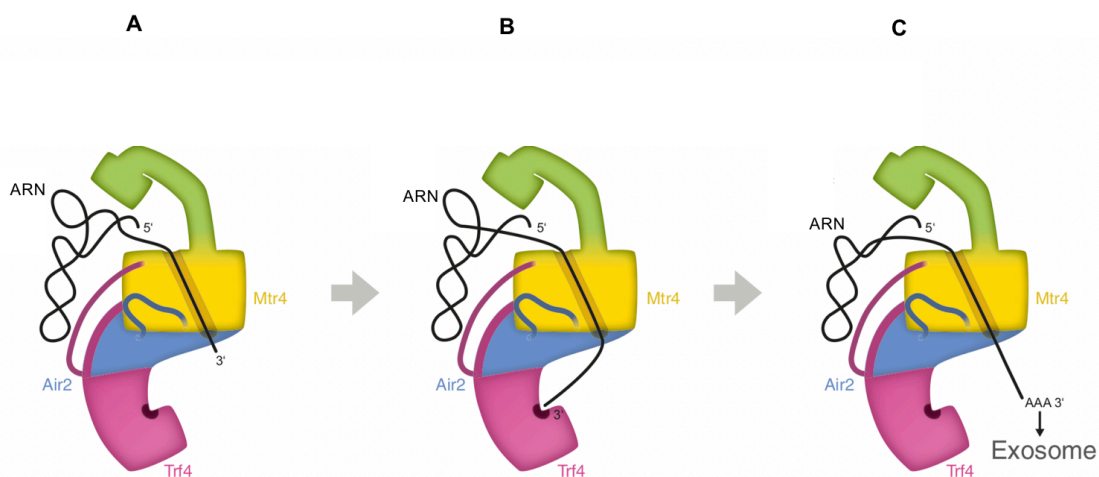
### Le complexe TRAMP :

Le complexe TRAMP (Trf4/5-Air1/Air2-Mtr4) est l'un des cofacteurs de l'exosome nucléaires le mieux caractérisé et est conservé au cours de l'évolution (revue : Schmidt and Butler, 2013; revue : Wong et al., 2015). Chez *S. cerevisiae*, le complexe TRAMP est un complexe hétérotrimérique constitué d'une poly(A) polymérase (Trf4 ou Trf5) ; d'une protéine de liaison de l'ARN (Air1p ou Air2p) ; et de l'hélicase Mtr4 (Houseley and Tollervey, 2006; LaCava et al., 2005; Losh et al., 2015) (figure 15). Trf4 et Trf5 ne contiennent pas de domaine de liaison à l'ARN contrairement aux autres poly(A) polymérases. Ainsi la liaison de ces protéines à leurs substrats se fait par l'intermédiaire d'une protéine accessoire, Air1 ou Air2. L'hélicase Mtr4 est capable de transloquer par l'ARN dans le sens 3'-5', ce qui lui permet d'éliminer des structures secondaires présentes dans l'ARN et permettre ainsi la polyadénylation de ces ARNs par Trf4/5. Plusieurs études ont montré que le complexe TRAMP est nécessaire d'une part à la polyadénylation des ARN non-codants cibles de la voie NNS tels que les précurseurs des SnoARNs (LaCava et al., 2005; Vanacova and Stefl, 2007) et les CUTs (Wyers et al., 2005). Enfin le complexe TRAMP cible aussi d'autres ARN non-codants tels que les ARN de transfert et les ARN ribosomiques (revue : Schmidt and Butler, 2013). L'ajout de la queue polyA par le complexe TRAMP stimule la dégradation des ARN aberrants par l'exosome.

Au cours de la terminaison de la transcription, le complexe NNS recrute le complexe TRAMP via une

interaction entre le domaine CID de Nrd1 et le NIM de Trf4 et il a été montré que cette interaction est importante pour une activité optimale de l'exosome nucléaire (Tudek et al., 2014).

Il existe trois formes distinctes du complexe TRAMP : Trf4-Air1-Mtr4 (TRAMP4-1); Trf4-Air2-Mtr4 (TRAMP4-2) et Trf5-Air1-Mtr4 (TRAMP5-1). Il a été proposé que ces trois formes distinctes permettent de conférer une spécificité de substrat au complexe. D'après le modèle proposé, TRAMP5-1 semble avoir une préférence pour l'ARN ribosomique 35S. En revanche, le TRAMP4-2 serait plus fortement associé aux gènes transcrits par l'ARNpol II, notamment certains ARN messagers dont le site de terminaison est proche du TSS mais aussi les CUTs, XUTs et SUTs. Ceci suggère l'implication de cette forme dans la dégradation des cibles classiques de la voie NNS. Quant au complexe TRAMP4-1, il semble être redondant avec le TRAMP4-2 mais pourrait avoir plus d'affinité pour les régions introniques des ARN messagers (Delan-Forino et al., 2020).



**Figure 15** (adaptée de : Falk et al., 2014) : Modèle du mécanisme d'action du complexe TRAMP.

La représentation schématique du complexe TRAMP réalisée à partir des informations structurales des travaux (Falk et al., 2014; Hamill et al., 2010; Jackson et al., 2010; Weir et al., 2010). L'hélicase Mtr4 est colorée en jaune (canal d'entrée de l'ARN en vert). La sous-unité Air2 est colorée en bleue et la polyA polymérase Trf4 en rose.

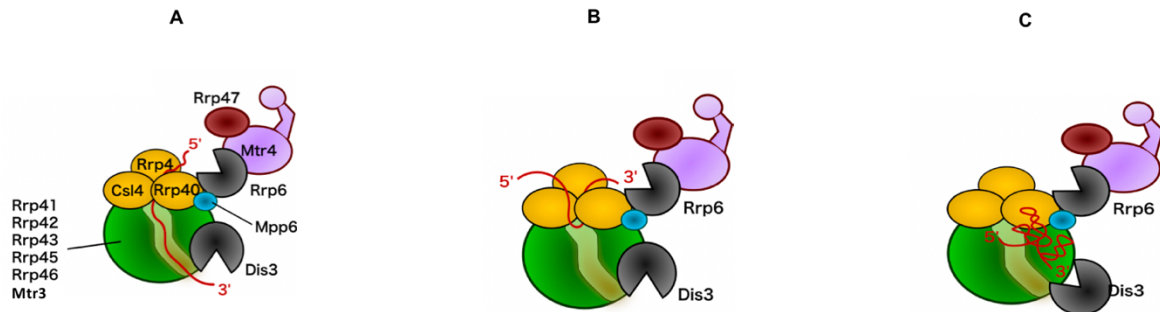
A – B : L'extrémité 3' de l'ARN s'enfile dans le canal de Mtr4 et atteint le site catalytique de Trf4.

C : Le détachement de l'ARN du site de polyadénylation de Trf4 permettrait à son extrémité 3' d'être accessible et dégradé par l'exosome.

### L'exosome nucléaire :

L'exosome est un complexe multiprotéique conservé au cours de l'évolution qui catalyse la maturation ou la dégradation de diverses classes d'ARNs (revue : Kilchert et al., 2016; revue : Ogami et al., 2018; revue : Zinder and Lima, 2017). Chez *S. cerevisiae*, il est composé de neuf sous-unités structurales (Rrp4, Rrp40, Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, Mtr3 et Csl4) appelées sous forme simplifiée Exo9 qui constitue un échafaudage permettant de recruter deux sous-unités catalytiques : Rrp6 et Rrp44 (Dis 3). Sa composition varie selon sa localisation qui peut être cytoplasmique ou nucléaire. La forme cytoplasmique est constituée de Exo9 associé à Rrp44 tandis que la forme nucléaire est formée de Exo9, Rrp44 et Rrp6 (revue : Zinder and Lima, 2017). Afin d'interagir avec le complexe TRAMP, Rrp6 interagit avec Rrp47 et l'ensemble constitue une plateforme de recrutement de Mtr4 (Schuch et al., 2014). La protéine accessoire Mpp6 se lie à la sous-unité Rrp40 d'Exo9 et stimule l'activité hélicase de Mtr4 (Falk et al., 2014; Wasmuth et al., 2017). Cette activité est nécessaire pour dérouler les structures secondaires formées à l'extrémité 3' des ARNs de manière à avoir un

ARN simple brin qui peut alors s'enfiler dans le canal de l'exosome dans une orientation 3'-5' (revue : Schneider and Tollervy, 2013). Rrp44 dégrade les ARNs qui s'enfilent par le canal central (figure 16A) tandis que Rrp6 dégrade les ARNs qui passent par l'anneau de la coiffe (figure 16B et Wasmuth et al., 2017, 2014; Zinder et al., 2016). En outre, il existe un autre site par lequel l'ARN peut accéder au site actif de Rrp6 (Liu et al., 2014) (figure 16C).



**Figure 16** (adaptée : Ogami et al., 2018) : Structure schématique de l'exosome eucaryote avec les chemins d'accès possibles des substrats vers les sous-unités catalytiques Rrp6 et Dis3

A : L'ARN pénètre dans le canal central et atteint le site actif de Dis3 ; B : L'ARN traverse un hétérotrimère (Rrp4, Csl4 et Rrp40) et atteint directement Rrp6 ; C : L'ARN contourne le canal central et atteint directement Dis3.

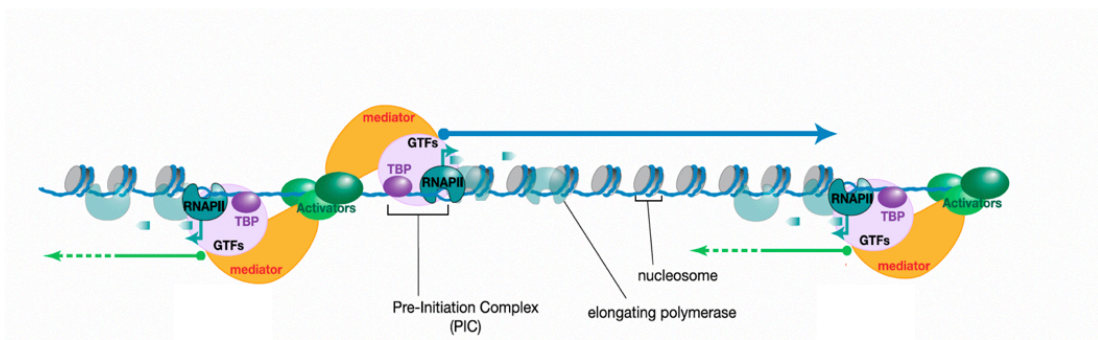
## 1.4 La transcription cachée ou pervasive :

### 1.4.1 Découverte, origine et contrôle de la transcription pervasive :

Les révolutions technologiques notamment avec le séquençage à haut débit au cours de ces vingt dernières années ont permis de révéler que la transcription n'est pas limitée aux unités codant les protéines et autres ARNs fonctionnels tels que les ARNr, ARNt, etc (1-2% du génome total). En réalité, elle concerne une fraction beaucoup plus importante de notre patrimoine génétique que l'on considèrerait transcriptionnellement inactive. Ainsi, des études de grande envergure réalisées chez l'homme ont permis de montrer qu'environ 75% du génome humain est transcrit (Djebali et al., 2012) tandis que chez *S. cerevisiae*, la fraction transcrite représenterait environ 85% du génome total (David et al., 2006). Cette transcription a été appelée la transcription cachée ou pervasive car la plupart des ARNs ainsi générés sont rapidement dégradés par des nucléases. Elle est conservée chez la plupart des eucaryotes et serait une source importante d'ARN non-codant généré par l'ARNpol II à partir des régions inter-géniques, au niveau des promoteurs et des terminateurs. L'existence de la transcription pervasive a été aussi mise en évidence chez la bactérie et joue un rôle important dans la régulation de l'expression des gènes (revue : Lybecker et al., 2014 et Wade and Grainger, 2014). Plusieurs études ont montré qu'une fraction importante de cette transcription non-codante provient de la bidirectionnalité des promoteurs générant ainsi des ARNs non-codants divergents de l'ARNm (Neil et al., 2009; Preker et al., 2008; Xu et al., 2009). En outre, il a été révélé que les gènes codants des protéines et des gènes générant des ARN non-codants utilisent des PICs distincts pour l'initiation de la transcription (Rhee and Pugh, 2012)

La figure 17 illustre des unités de transcription de l'ARN messagers typiques des eucaryotes avec l'organisation des facteurs de transcription des gènes et des ARN non-codants.





**Figure 17** (adaptée de : Torben Heick Jensen et al., 2013) : Schéma illustrant la transcription pervasive à partir des promoteurs de gènes codant des protéines. La production d'ARN non-codant en antisens provient d'un PIC distinct ayant la même composition que celui qui pilote la transcription de l'ARNm. La flèche bleue indique l'ARN messager tandis que les flèches vertes correspondent aux ARN non-codants.

Comme mentionné ci-dessus, la vaste majorité de ces ARN non-codants sont instables. Ainsi, plusieurs laboratoires dont le nôtre ont contribué à la découverte de différentes classes d'ARN non-codants chez la levure. Différents mutants dans la machinerie de dégradation ont permis de les classer en différents groupes. Ainsi, la mutation de la sous-unité catalytique de l'exosome nucléaire Rrp6 a permis de révéler environ 1500 ARN non-codants qui sont indétectables dans une souche sauvage appelée CUT (« Cryptique unstables transcripts »)(Neil et al., 2009; Xu et al., 2009). Ces ARN non-codants sont cappés, polyadénylés et relativement courts, avec une longueur moyenne de 200 à 500 pb, et contiennent des extrémités 3' hétérogènes (Wyers et al., 2005). Suite à l'inactivation simultanée de Rrp6 et de l'autre sous-unité catalytique de l'exosome, Dis3, plusieurs centaines de nouveaux CUTs ont été identifiés (Gudipati et al., 2012). En plus des CUTs, une classe d'ARN non-codant partiellement sensible à l'inactivation de l'exosome nucléaire appelé SUTs (« Stable unannotated transcripts ») furent identifiés. A la différence des CUTs, les SUTs sont suffisamment stables pour être détectables dans une souche sauvage (Xu et al., 2009). Une fraction importante des SUTs est terminée par la voie CPF-CF et leur dégradation est effectuée en grande partie dans le cytoplasme par le « Nonsense mediated decay » (NMD). Une autre fraction est effectuée par la voie NNS (Malabat et al., 2015; Neil et al., 2009; Schulz et al., 2013; Xu et al., 2009). Ultérieurement, il a été montré que la délétion de Xrn1, une exoribonucléase cytoplasmique, conduit à la stabilisation d'environ 1600 ARN non-codants appelés XUTs (« Xrn1-sensitive unstables transcripts ») dont 66% sont orientés en antisens des gènes (van Dijk et al., 2011). Les XUTs se chevauchent souvent avec des CUTs et des SUTs et correspondent à une fraction des CUTs et SUTs qui ont échappé à la voie terminaison de la transcription couplée à la dégradation nucléaire et sont ainsi dégradés dans le cytoplasme.

Il est important de mentionner que la plupart des études à grande échelle ont été réalisées dans un nombre limité de conditions physiologiques (typiquement croissance exponentielle en milieu riche) et que chaque nouvelle étude en conditions non-standard, par exemple quiescence, stress, etc, a permis de dévoiler des nouveaux ARN non-codants (Nadal-Ribelles et al., 2014; Nevers et al., 2018). Ainsi, une fraction importante de la transcription cachée reste alors à caractériser.

Dans l'ensemble, la transcription pervasive présente un risque potentiel pour la cellule qu'il convient de contrôler car elle peut s'interférer avec la transcription normale des gènes. Chez *Saccharomyces cerevisiae*, l'acteur majeur dans le contrôle de la transcription pervasive est le complexe Nrd1-Nab3-Sen1 (NNS), qui induit l'arrêt précoce de la transcription non-codante et favorise la dégradation des ARNs générés par l'exosome nucléaire par les mécanismes décrits ci-dessus.

### 1.4.2 Fonction de la transcription pervasive :

Le caractère fonctionnel de la transcription pervasive a fait débat au sein de la communauté scientifique. Certains auteurs proposent que ces ARN non-codants pourraient au cours de l'évolution générer de protéines dotées de nouvelles fonctions (Van Oss and Carvunis, 2019). En outre, dans un nombre croissant de cas, la transcription non-codante s'est avérée jouer un rôle important dans la régulation de l'expression des gènes. Cette régulation peut s'opérer en *cis* ou en *trans*. En *trans*, c'est-à-dire que la molécule d'ARN non-codant qui possède potentiellement une fonction régulatrice, une fonction qui sera perdue par la dégradation de l'ARN non-codant. Chez *S. cerevisiae*, le premier et probablement le seul exemple de régulation en *trans* décrit dans la littérature est la régulation du retrotransposon Ty1 par un ARN non-codant en antisens. Berretta et ses collègues ont montré que l'exoribonucléase Xrn1 régule et maintient l'expression du gène Ty1 en empêchant l'accumulation de l'ARN antisens. La stabilisation de l'ARN en antisens favorise des modifications épigénétiques conduisant à la répression de Ty1 (Berretta et al., 2008). En *cis*, la transcription pervasive a elle-même un impact sur la régulation du gène en aval ou en antisens que l'ARN non-codant produit soit dégradé ou non. Chez *S. cerevisiae*, la quasi totalité des régulations par la transcription pervasive décrits s'opèrent par des mécanismes en *cis* (voir tableau 1).

Le premier exemple de ce type de régulation fût révélé par Martens et ses collaborateurs en 2004. Il s'agit du gène *SER3* régulé par un ARN non-codant en amont appelé SRG1 (« SER3 regulatory gene 1 ») (Martens et al., 2004). Le gène *SER3* code pour une protéine catalysant une étape clé de la biosynthèse de la sérine. Cette régulation impliquant la transcription de l'ARN non-codant SRG1 permet de moduler la réponse cellulaire en fonction de la disponibilité de cet acide aminé essentiel pour la croissance cellulaire. L'expression de *SER3* est activée en absence de sérine et est réprimée dans un milieu riche en sérine.

En présence de concentrations non-limitantes de sérine, le facteur de transcription Cha4 recrute le complexe de remodelage de la chromatine SWI/SNF et le complexe SAGA à la région promotrice du gène *SRG1*. Ces facteurs réunis activent la transcription de *SRG1* qui peut s'étendre jusqu'au nucléotide +1 du gène *SER3* et empêche ainsi la liaison des activateurs de la transcription à la région promotrice du gène *SER3* (Martens, 2005) conduisant à sa répression par interférence transcriptionnelle. Dans la continuité de la même étude, un autre travail plus récent a montré que ce mécanisme d'interférence transcriptionnelle est médié par l'accumulation des nucléosomes au niveau de la région promotrice du gène *SER3* lors de la transcription de *SRG1*. Ce processus implique l'action de Spt6 et Spt16, deux chaperonnes d'histones qui facilitent le déplacement des nucléosomes au cours de la transcription (Hainer et al., 2011).

En absence de sérine, le facteur de transcription Cha4 n'interagit pas avec les complexes SWI/SNF et SAGA et donc l'activation de *SRG1* ne s'effectue pas. L'organisation des nucléosomes autour de la région promotrice de *SER3* est donc favorable à son expression.

En plus, il a été montré que l'ARN non-codant SRG1 contient deux sites de terminaison consécutifs : le premier terminateur dépend de la voie NNS et le second, de la voie CPF-CF. Si *SRG1* est fortement transcrit, l'une des voies de terminaison peut être saturée, donc une seconde voie de terminaison peut être un moyen supplémentaire pour terminer la transcription de *SRG1* plus efficacement et empêcher la production d'ARN chimérique SRG1-SER3. En effet, la déplétion de Nrd1 conduit à l'accumulation de cette forme chimérique (Thiebaut et al., 2006) qui pourrait éventuellement donner lieu à une forme potentiellement toxique de la protéine Ser3.

ARN non-codant	Gène cible	Processus physiologique	Facteurs épigénétiques impliqués dans la régulation	Références
asPHO5	<i>PHO5</i>	Métabolisme du phosphate	?	(Castelnuovo et al., 2014a)
IRT1	<i>IME1</i>	Méiose	Set2 et Set3	(van Werven et al., 2012)
SRG1	<i>SER3</i>	Métabolisme de la sérine	?	(Hainer et al., 2011; Martens, 2005; Martens et al., 2004)
ZRR1/ZRR2	<i>ADH1/ADH3</i>	Métabolisme du zinc	?	(Amanda J Bird et al., 2006)
ARNnc <i>GAL10</i>	<i>GAL10</i> <i>GAL1</i>	Métabolisme du galactose	Set2, RpdS3	(Houseley et al., 2008)
asCDC28	<i>CDC28</i>	Stress osmotique	Ssu72	(Nadal-Ribelles et al., 2014b)
ARNnc <i>ASP3</i>	<i>ASP3</i>	Métabolisme de l'azote	?	(Huang et al., 2010)
RME2	<i>IME4</i>	Méiose	?	(Hongay et al., 2006)
RME3	<i>ZIP2</i>	Méiose	?	(Gelfand et al., 2011)
ICR1	<i>FLO11</i>	Adhésion cellule – cellule	Rpd3L	(Bumgarner et al., 2009)
PWR1	<i>FLO11</i>	Adhésion cellule – cellule	Rpd3L	(Bumgarner et al., 2009)
ARNnc Pho	Gène <i>HO</i>	Différenciation sexuelle	?	(Yu et al., 2016)
asPHO84	<i>PHO84</i>	Métabolisme du phosphate	Hda1, 2 et 3	(Camblong et al., 2007a; Castelnuovo et al., 2013a, 2014a)

**Tableau 1** : Récapitulatif des principaux ARN non-codants dont la transcription joue un rôle dans la régulation de l'expression d'un gène voisin. Les ARN non-codants transcrits en antisens des gènes régulés sont nommés "as".

Un deuxième exemple concerne les gènes *ADH1* et *ADH3*, qui sont co-réprimés en conditions de limitation de zinc. Cette régulation implique Zap1, le facteur de transcription qui permet de moduler l'expression d'un certain nombre de gènes en réponse à la disponibilité de zinc.

*ADH1* code pour l'alcool déshydrogénase, une enzyme dépendante du zinc jouant un rôle clé dans la fermentation. Afin de s'adapter à la limitation de zinc, la répression des gènes codant des protéines consommatrices de zinc fait partie du mécanisme d'adaptation. En situation de carence en zinc, Zap1 se lie en amont du site de liaison de l'activateur de la transcription de *ADH1*, Rap1, et active la transcription d'un ARN non-codant appelé ZRR1 ("Zinc Regulated RNA 1") à partir de la région intergénique. La transcription de ZRR1 empêche le recrutement de l'activateur de la transcription Rap1 et par conséquent conduit à la répression de *ADH1*. Dans la continuité du même travail, les auteurs ont montré que *ADH3*, codant pour une autre alcool déshydrogénase mitochondriale, est régulée de manière similaire à *ADH1* (Bird et al., 2006).

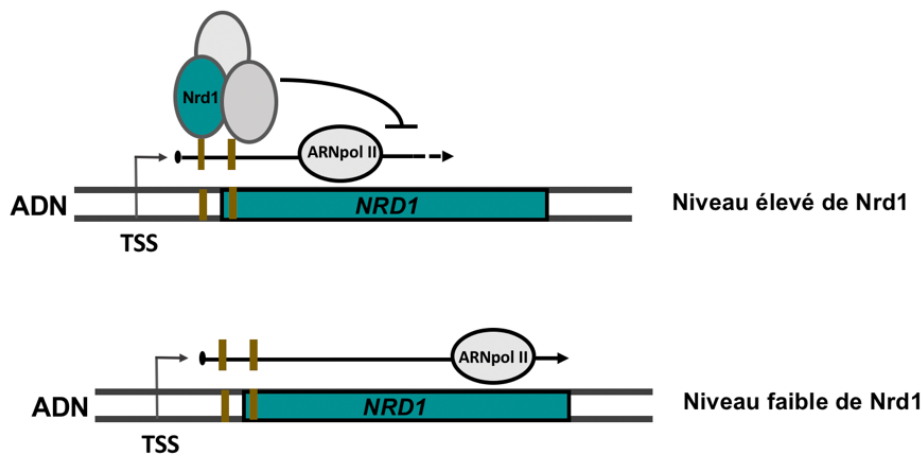
Enfin, un autre exemple de régulation par la transcription non-codante concerne le gène *PHO84* dont la régulation a été caractérisée en détail par le laboratoire de Françoise Stutz (Université de Genève). Ce laboratoire a tout d'abord montré que la transcription du gène *PHO84* est réprimé par la transcription de l'ARN non-codant en antisens (asPHO84) en absence de Rrp6, indiquant un rôle de l'exosome dans la régulation de ce gène. L'absence de Rrp6 favorise l'accumulation et l'extension de l'ARN de l'asPHO84 et provoque le recrutement des histones desacétyltransférases au niveau du promoteur du gène *PHO84* afin d'induire sa répression (Camblong et al., 2007). Afin d'élucider d'avantage ce mécanisme de régulation, les auteurs ont entrepris une expérience d'hybridation *in situ* (SM-FISH) afin d'analyser concomitamment l'expression du sens et l'antisens. Ces expériences ont confirmé que la transcription du sens et de l'antisens est mutuellement exclusive suggérant un mécanisme de régulation de type interrupteur. De plus, ces données ont permis de révéler que la délétion de Rrp6 affecte négativement la terminaison précoce de la transcription de l'asPHO84 par le complexe NNS et cela conduit à une augmentation de la transcription antisens à travers le promoteur de *PHO84* et, donc, une répression de l'expression de *PHO84* par interférence transcriptionnelle. Cette répression dépendrait de l'histone méthyltransférase Set1 (Castelnuovo et al., 2013). Une étude réalisée au sein de notre laboratoire a montré que le mécanisme par lequel la délétion de RRP6 empêche une terminaison efficace de la transcription de certains ARN non-codants de la voie NNS. Nrd1 et Nab3 sont titrées par la non-dégradation et accumulation des ARNs cibles, se trouvent ainsi piégés et donc incapables de s'associer à d'autres ARN naissants et de promouvoir la terminaison de la transcription (Villa et al., 2020)

Parallèlement, le laboratoire de Françoise Stutz a réalisé des études transcriptomiques afin d'analyser le rôle du complexe NNS et les modifications épigénétiques dans la répression génique. Cette étude a permis d'identifier un certain nombre de gènes sensibles à l'interférence transcriptionnelle et ces gènes ont pu être catégorisés. Dans la catégorie la plus distinctive, les gènes sensibles à l'interférence transcriptionnelle dépendante de la délétion de *RRP6* sont généralement ceux qui ont un ARN non codant en antisens qui sont terminés au moins partiellement terminés par la voie NNS (Castelnuovo et al., 2014). Ces résultats impliquent que des mécanismes modulant négativement l'action du complexe NNS pourraient potentiellement réguler l'expression d'un certain nombre de gènes par l'interférence transcriptionnelle.

## 1.5 Rôle de la voie NNS dans la régulation de l'expression des gènes

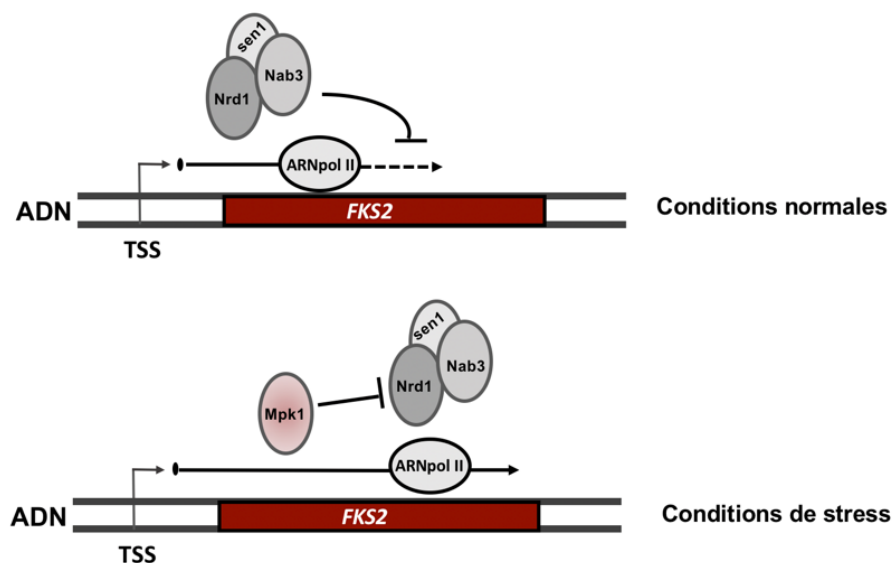
Il a été montré que le complexe NNS joue un rôle dans la régulation de l'expression des gènes par la terminaison précoce ou atténuation transcriptionnelle, qui conduit à la répression de la transcription. Le mécanisme d'atténuation a été pour la première fois découvert chez les bactéries. Elle s'appuie sur la présence d'une séquence particulière d'ARN au niveau de la région 5' UTR du gène régulé, qui a la capacité de se replier pour former un terminateur de la transcription dans certaines conditions environnementales (revue : Naville and Gautheret, 2009). En revanche chez la levure, la régulation par atténuation apparaît moins fréquente et semble être généralement constitutive.

L'un des exemples les mieux étudiés de régulation par atténuation découvert chez *S. cerevisiae* est l'autorégulation de *NRD1* (Arigo et al., 2006b). Le locus *NRD1* contient des sites de liaison pour Nrd1 et Nab3 dans la région 5' du gène, et la terminaison dépendante du complexe NNS se produit avec une efficacité qui dépend du niveau de la protéine Nrd1, établissant ainsi une boucle de rétroaction négative (figure 18).



**Figure 18 :** Mécanisme de régulation de l'expression de Nrd1 par terminaison prématurée. Lorsque le niveau intracellulaire de Nrd1 est élevé, Nrd1 est recruté au niveau de son propre locus afin d'induire la terminaison prématurée. En revanche, lorsque son niveau intracellulaire est faible, le gène est transcrit et terminé par la voie CPF-CF générant ainsi un ARN messager traduit en protéine. Les boîtes marrons désignent les sites de liaisons de Nrd1 et Nab3.

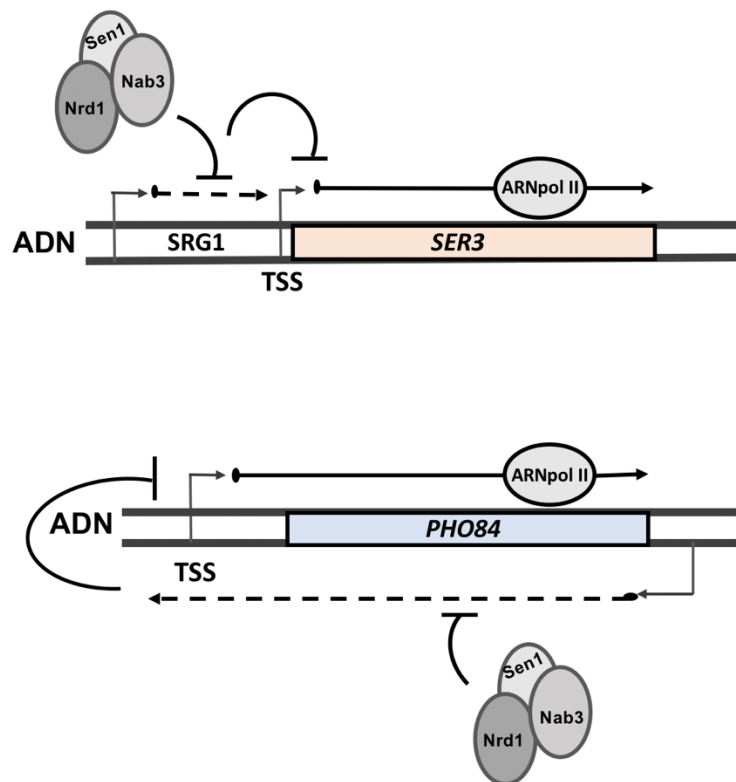
Un autre mécanisme de régulation par atténuation a été proposé aussi pour le contrôle de l'expression du gène *FKS2*, qui code pour une 1,3-bêta-glucan synthase, une protéine impliquée dans la synthèse d'un composant structural majeur de la paroi cellulaire. En condition normale de croissance, l'expression du gène *FKS2* est réprimée et la terminaison précoce de la transcription par le complexe NNS contribue aux niveaux globaux de répression. En condition de stress de la paroi cellulaire, comme par exemple une température élevée ou la présence de molécules qui compromettent l'intégrité de la paroi, la cellule a besoin de renforcer sa paroi et active la voie de signalisation dépendante de MAPK (mitogen-activated protein kinase). Mpk1 induit la transcription du gène *FKS2* en recrutant les facteurs de transcription au niveau du promoteur de ce gène. En parallèle, l'association de Mpk1 avec Paf1 dans la région codante du gène *FKS2* semblerait interférer avec le recrutement de Sen1 empêchant ainsi la terminaison précoce, ce qui permet une activation complète de l'expression de *FKS2* (Kim and Levin, 2011) (figure 19).



**Figure 19 :** Modèle de régulation de l'expression de *FKS2* par la terminaison précoce dépendante du complexe NNS.

De plus, une étude transcriptomique a permis de montrer qu'environ 32 gènes sont potentiellement affectés dans la terminaison précoce en réponse à la déplétion de Nrd1 (Schulz et al., 2013). Ceci montre que la régulation de l'expression des gènes par la terminaison précoce est loin de se limiter à ces deux exemples même si ce mode de régulation reste globalement rare chez la levure, au moins dans les conditions standard de croissance.

Le complexe NNS peut aussi participer à la régulation de l'expression des gènes indirectement c'est-à-dire par le biais de la transcription d'un ARN non codant. Deux exemples de ce type de régulation seraient la répression du gène *SER3* par la transcription du gène non-codant en amont *SRG1* et la répression du gène *PHO84* par la transcription d'un antisens (décrits dans la section précédente). Dans les deux cas, la transcription non-codante est au moins partiellement terminée par le complexe NNS et, donc, en modulant l'efficacité de la terminaison par NNS, les niveaux de répression par la transcription de ces ARN non-codants pourraient être modifiés (figure 20).



**Figure 20** : représentation de deux exemples dans lequel la terminaison de la transcription non-codante par le complexe NNS peut jouer un rôle dans la régulation de gènes. Le complexe NNS cible des ARN non-codants dont la transcription réprime l'expression du gène associée par interférence transcriptionnelle.

Des analyses transcriptomiques ont permis de révéler l'existence d'environ 1500 ARN non-codants dont la terminaison est dépendante du complexe NNS (Schulz et al., 2013). En plus, de défauts de terminaison dans 200 à 300 de ces ARN non-codants suite à la déplétion de la protéine Nrd1 seraient associés directement à la dérégulation des gènes qui se trouvent soit en aval, soit en antisens (Gill et al., 2020; Schulz et al., 2013). Ces résultats mettent en évidence le potentiel du rôle régulateur de la terminaison de la transcription non-codante. Nous pouvons donc émettre l'hypothèse selon laquelle l'action du complexe NNS pourrait être soumise à des régulations afin de moduler l'expression des gènes.



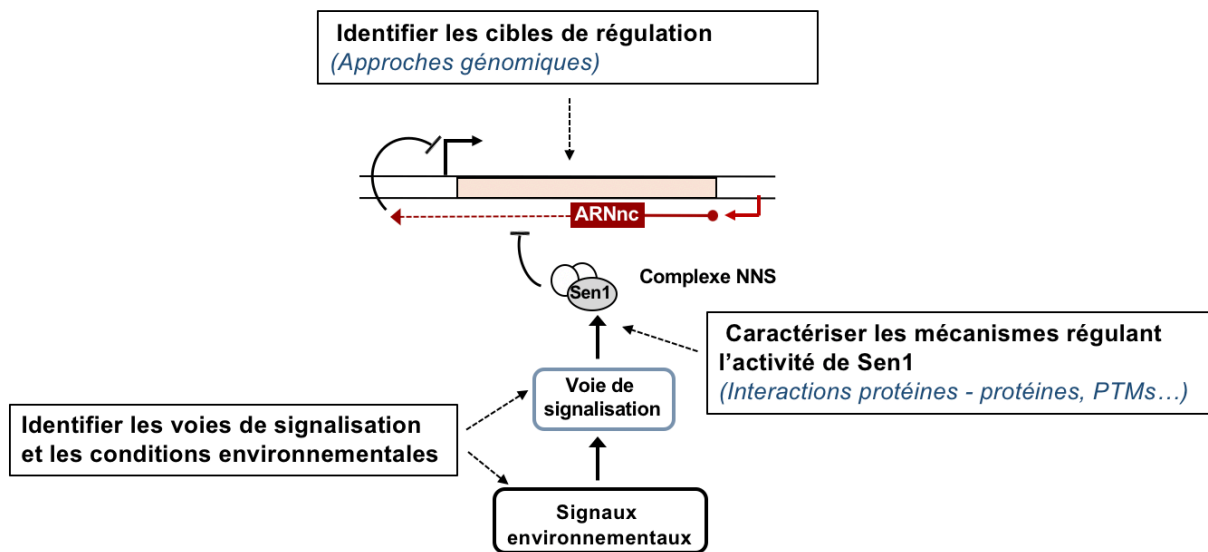
## 2 Objectifs du projet et stratégie expérimentale



## Objectifs du projet et stratégie expérimentale

L'ensemble des données évoquées dans l'introduction suggèrent que la terminaison de la transcription non-codante est potentiellement une étape importante dans la régulation des gènes. Durant ma thèse, j'ai exploré l'existence possible d'un mécanisme de régulation de l'expression des gènes qui s'opère en modulant l'efficacité de la terminaison de la transcription non-codante en réponse aux changements environnementaux.

Comme décrit au cours des sections précédentes, chez *S. cerevisiae*, le complexe NNS est la machinerie qui assure la terminaison d'une grande fraction de la transcription non-codante. Au sein de ce complexe, l'hélicase Sen1 la sous unité catalytique est le facteur responsable de la dissociation du complexe d'élongation. Notre hypothèse est que la plupart des régulations pourraient s'opérer sur Sen1. Pour tester cette hypothèse, j'ai tenté de combiner des approches génétiques, protéomiques et génomiques pour : **i)** caractériser les mécanismes qui contrôlent l'activité de Sen1 par des interactions intermoléculaires et /ou par des modifications post-traductionnelles (PTMs); **ii)** comprendre comment ces mécanismes peuvent avoir un impact sur l'efficacité de la terminaison de la transcription non codante et contribuer à la régulation l'expression des gènes ; **iii)** enfin, identifier les conditions physiologiques dans lesquelles la protéine Sen1 serait régulée ainsi que les voies de signalisation qui ciblent Sen1 pour moduler l'efficacité de la terminaison de la transcription non-codante afin de réguler l'expression des gènes (figure 21).



**Figure 21** : Schéma récapitulatif des objectifs de ce projet et la stratégie expérimentale exploitée pour aborder ces questions. PTMs indique « post-translational modifications »

### 3 Résultats et discussions

### 3.1 La terminaison modulée de la transcription non-codante participe à la régulation de l'expression des gènes

# **Modulated termination of non-coding transcription partakes in the regulation of gene expression**

Haidara, N<sup>1,2</sup> and O. Porrua<sup>1\*</sup>

<sup>1</sup> Université de Paris, CNRS, Institut Jacques Monod, Paris, France.

<sup>2</sup> Université Paris-Saclay, Gif sur Yvette, France.

\*Correspondence should be addressed to: odil.porrua@ijm.fr

## **ABSTRACT**

Pervasive transcription is a universal phenomenon leading to the production of a plethora of non-coding RNAs. If left uncontrolled, pervasive transcription can be harmful for genome expression and stability. However, non-coding transcription can also play important regulatory roles, for instance by promoting the repression of specific genes by a mechanism of transcriptional interference. The efficiency of transcription termination can strongly influence the regulatory capacity of non-coding transcription events. Yet very little is known about the mechanisms modulating the termination of non-coding transcription in response to environmental cues.

Here, we address this question by investigating the mechanisms that regulate the activity of the main actor in termination of non-coding transcription in budding yeast, the helicase Sen1. We identify a phosphorylation at a conserved threonine of the catalytic domain of Sen1 and we provide evidence that phosphorylation at this site reduces the efficiency of Sen1-mediated termination. Interestingly, we find that this phosphorylation impairs termination at an unannotated non-coding gene, thus repressing the expression of a downstream gene encoding the master regulator of Zn homeostasis, Zap1. Consequently, many additional genes exhibit an expression pattern mimicking conditions of Zn excess.

Our findings provide a novel paradigm of gene regulatory mechanism relying on the direct modulation of non-coding transcription termination.

## INTRODUCTION

One of the most fascinating discoveries over the last years on the field of gene expression has been the finding that transcription is not restricted to annotated genes but actually extends to the vast majority of the genome. This phenomenon, dubbed pervasive transcription, has been revealed in a multitude of organisms from bacteria to humans (reviewed in Torben Heick Jensen et al., 2013). Pervasive transcripts are often rapidly degraded, and therefore most of them remain invisible unless RNA degradation is prevented. In general, pervasive transcripts originate from nucleosome-free regions, either from cryptic promoters or from *bona fide* gene promoters that generate transcription divergently from the controlled gene with a certain frequency. Indeed, several studies have concluded that the vast majority of eukaryotic promoters are intrinsically bidirectional (H. Neil et al., 2009; Preker et al., 2008; Schulz et al., 2013; Z. Xu et al., 2009).

Pervasive transcription is potentially detrimental to cell homeostasis because, if left uncontrolled, it can invade and interfere with genomic regions involved in gene regulation, DNA replication or repair (Candelli et al., 2018; Schulz et al., 2013; Soudet et al., 2018). Therefore, all organisms studied to date have evolved different mechanisms to limit the negative consequences of pervasive transcription. These mechanisms often rely on early termination of non-coding transcription and ncRNA degradation (reviewed in Torben Heick Jensen et al., 2013).

The mechanisms responsible for the control of pervasive transcription have been particularly well characterized in budding yeast. In this organism, the so-called NNS (Nrd1-Nab3-Sen1) complex induces early termination of pervasive transcription and promotes the degradation of the resulting ncRNAs by the nuclear exosome (for review, see Porrua and Libri, 2015). Nrd1 and Nab3 are RNA-binding proteins that recognize specific sequences that are enriched in the target ncRNAs (Hobor et al., 2011; Lunde et al., 2011; Porrua et al., 2012; Schulz et al., 2013; Wlotzka et al., 2011), while the conserved helicase Sen1 is the catalytic subunit of the complex (Porrua and Libri, 2013). The association of Sen1 with Nrd1 and Nab3 enhances Sen1 recruitment to ncRNAs, albeit is not a strict requirement for Sen1 loading on these RNAs (Han et al., 2020). Finally, Sen1 uses the energy of ATP hydrolysis to translocate along the nascent RNA and induce the release of RNAPII from the DNA (Han et al., 2017; Porrua and Libri, 2013).

The widespread character of pervasive transcription has raised the question of its possible biological role. Unlike metazoans, where ncRNAs can play multiple regulatory roles, in budding yeast it is the act of transcription rather than the ncRNA produced which has proved to be functional. Non-coding transcription can interfere with the expression of genes that are located just downstream or in the antisense orientation when it invades the promoter region of these genes. The mechanisms at play typically involve the deposition of repressive chromatin marks at the promoter of regulated genes (Castelnuovo et al., 2014; Gill et al., 2020; Huber et al., 2016; Nevers et al., 2018). There is a growing number of well-characterized examples where non-coding transcription mediates the regulation of gene expression in response to environmental cues. These include genes involved in amino acid biosynthesis, meiosis, stress, etc (Martens et al., 2005; Nadal-Ribelles et al., 2014; van Werven et al., 2012). In many cases, the ncRNA is expressed from a promoter that is only active under specific conditions (i.e. conditions in which the regulated protein-coding gene needs to be repressed). Paradigmatic examples of this mode of regulation are the *ADH1* and *ADH3* genes, which encode two abundant Zn-dependent alcohol dehydrogenases whose expression is repressed in response to Zn limitation by the upstream non-coding genes *ZRR1* and *ZRR2*, respectively. The expression of *ZRR1* and *ZRR2* is activated by the master regulator of Zn homeostasis, Zap1, and transcription through the promoter regions of *ADH1* and *ADH3* interferes with promoter recognition by the corresponding transcriptional activators (Amanda J. Bird et al., 2006).

More than 1500 non-coding genes expressed under standard growth conditions are targeted by the NNS-dependent pathway for early transcription termination coupled to ncRNA degradation. Importantly, in many of these cases depletion of the Nrd1 component of the NNS-complex abrogates termination by this pathway and provokes misregulation of the neighboring genes (Gill et al., 2020; Schulz et al., 2013), indicating that the efficiency of non-coding transcription termination is a potentially relevant parameter in the regulation of gene expression. However, thus far it remains unclear whether the activity of the NNS-complex on non-coding genes can be modulated in response to specific environmental signals to promote gene regulation. To address this question, we have focused on the mechanisms that control the activity of the catalytic subunit of the NNS-complex, the helicase Sen1. We have identified a phosphorylation at a conserved residue of Sen1 catalytic

domain and we have shown that phosphorylation at this site (T1623) impairs the interaction of Sen1 with the RNA and reduces the efficiency of transcription termination both *in vitro* and *in vivo*. Genome-wide transcription analyses revealed that a T1623 phospho-mimetic mutation provoked repression of the *ZAP1* gene via transcriptional interference exerted by an upstream non-annotated gene that we have dubbed *ZRN1* for Zap1-Repressor Non-coding gene 1. This resulted in an alteration in the expression of many Zap1-dependent genes involved in Zn homeostasis and an overall transcription profile reminding conditions of Zn excess, where *ZAP1* is naturally repressed. Our data suggested that phosphorylation at Sen1 T1623 also promotes coordinated repression of at least one additional gene by transcriptional interference to further contribute to the regulation of Zn intracellular levels.

Our results support a model whereby the catalytic domain of Sen1 would be phosphorylated under specific conditions to modulate its activity and regulate the expression of genes involved in Zn homeostasis.

## RESULTS

### ***Identification of a phosphorylation at a conserved threonine of Sen1 helicase domain***

To identify possible mechanisms of regulation of Sen1-mediated transcription termination we undertook an analysis of Sen1 post-translational modifications. To this end, we immunoprecipitated a C-terminally TAP-tagged version of Sen1 expressed at the endogenous locus and performed mass spectrometry analyses on Sen1 immunoprecipitates on 5 biological replicates (table S1). We identified 7 phosphosites, among which 6 are novel, spread all along Sen1 functional domains (figure 1A). Sen1 possesses an essential superfamily 1 helicase domain (aa 1095-1905) flanked by a large N-terminal domain (aa 1-975) and a C-terminal intrinsically-disordered region (aa 1906-2231), both involved in protein-protein interactions that are important for Sen1 function (Finkel et al., 2010; Han et al., 2020; Ursic et al., 2004). We considered particularly interesting the phosphosites located at the helicase domain (HD), since we have previously shown that this domain is necessary and sufficient for transcription termination *in vitro* (Han et al., 2017). Among them, phosphorylation of T1623 was detected in two independent experiments and the specific location of this residue made its phosphorylation a good candidate for mediating the regulation of Sen1 catalytic activity. Indeed, a comparison of the structure of Sen1 with that of the closest paralogue of Sen1, Upf1, in complex with the RNA (figure 1B) indicated that T1623 is placed at the predicted RNA-binding region of Sen1. This region could be inferred from the structural similarity between both proteins as well as from our previous mutational analysis confirming that conserved amino acids located at Upf1 RNA-binding surface are also involved in Sen1 interaction with the RNA (Leonaité et al., 2017). Because the predicted RNA-binding region of Sen1 is rather positively charged, we hypothesized that phosphorylation of T1623, by introducing a negative charge, could interfere with Sen1 RNA-binding.

Alignment of Sen1 proteins from diverse organisms revealed that the T1623 residue is highly conserved from budding yeast to humans (figure 1C), further supporting the idea that this threonine is important for Sen1 activity. Interestingly, in Upf1 the equivalent amino acid is a valine (V605, see figure 1B), which differs from a threonine exclusively in the presence of a methyl instead of the hydroxyl group of the threonine. This suggests that a threonine at that particular position might have been selected in Sen1 proteins throughout evolution either to confer specific properties to Sen1 HD or to confer the capacity to modulate Sen1 catalytic activity by phosphorylation.

### ***Phosphorylation of Sen1 at T1623 partially inhibits Sen1 activity both in vivo and in vitro***

In order to assess the role of T1623 phosphorylation in the regulation of Sen1 function, we introduced mutations either to prevent (T to V) or to mimic (T to E) phosphorylation of this residue and we analyzed the effect of these mutations on cell growth (figure 2A). Mutation to valine did not impair cell growth at any of the temperatures tested. In contrast, the T1623E phospho-mimetic mutation induced a strong growth defect at 37°C, while growth at other temperatures was not significantly affected. This thermosensitive phenotype was not due to decreased stability of the protein at high temperature, since Sen1 protein levels remain similar in cells grown at 30°C and 37°C (figure S1). Because termination of non-coding transcription by Sen1 is essential for viability, this

results suggests that phosphorylation might affect negatively Sen1 capacity to promote transcription termination. To test this possibility, we analysed by northern blot the expression of two typical Sen1 targets, namely the CUT (cryptic unstable transcript) NEL025c and the snoRNA *SNR13* (figure 2B). CUTs represent one of the major classes of pervasive transcripts while snoRNAs are functional non-coding RNAs that depend on the NNS-pathway for transcription termination and 3' end processing (Arndt and Reines, 2015b). Importantly, the efficiency of transcription termination at both non-coding genes was identical in the *sen1T1623V* mutant compared to the wild-type (wt) strain, indicating that a T and a V are interchangeable at position 1623 for normal Sen1 function in transcription termination. On the contrary, in the *sen1T1623E* mutant we observed a mild accumulation of transcripts ~200-500 nt longer than in the wt, which is suggestive of delayed transcription termination. This accumulation was very moderate under permissive temperature (i.e. at 30°C) but was strongly exacerbated in cells grown for 2h at 37°C (figure S1B), consistent with the thermosensitive phenotype of the *sen1T1623E* mutant. Taken together, these results strongly suggest that phosphorylation of Sen1 T1623 modulates negatively the catalytic activity of Sen1 and, therefore, the efficiency of Sen1-dependent transcription termination.

To gain mechanistic insight into the impact of T1623 phosphorylation on Sen1 catalytic activity, we expressed and purified from *E. coli* the wt, the T1623V and the T1623E versions of Sen1 HD (figure S2) and we characterized them with various biochemical assays. As indicated above, the HD of Sen1 recapitulates *in vitro* the mechanisms of termination employed by the full-length protein (Han et al., 2017; Leonaitė et al., 2017). These mechanisms involve the interaction of Sen1 with the nascent transcript and Sen1 translocation along the RNA towards RNAPII, which strictly requires ATP-hydrolysis by Sen1 (Han et al., 2017; Leonaitė et al., 2017; Porrua and Libri, 2013). To assess whether phosphorylation of T1623 would affect the association of Sen1 with the RNA, as the location of this residue predicts, we performed electrophoretic mobility shift assays (EMSA) with the three purified versions of Sen1 HD and a fluorescently-labeled 40-mer RNA as the substrate (figure 2C and S2). These experiments were carried out at 30°C and 37°C. The presence of the T1623V mutation did not decrease the affinity of Sen1 for the RNA at any of the temperatures tested. However, we observed a mild decrease in the affinity of Sen1T1623E for the RNA compared to the wt at 30°C, which was significantly more pronounced at 37°C. Because the interaction of Sen1 with the RNA is required to activate Sen1 ATPase activity, we performed ATP hydrolysis assays using non-saturating concentrations of RNA to assess the RNA-dependent ATPase activity of the different Sen1 HD variants (figure 2D). Consistent with the results of EMSA experiments, we observed reduced levels of ATP hydrolysis by the T1623E but not the T1623V version of Sen1, compared to the wt at 30°C. Furthermore, this decrease in Sen1 ATPase activity was strongly exacerbated at 37°C. Taken together, these results strongly suggest that the phosphorylation of Sen1 T1623 destabilizes the interaction of Sen1 with the RNA and provide the molecular basis of the thermosensitive phenotype observed *in vivo*.

To assess whether impaired RNA-binding in the T1623E mutant results in decreased transcription termination efficiency *in vitro*, we employed a highly purified *in vitro* transcription termination (IVTT) assay that we have previously developed (Porrua and Libri, 2015, see methods for details). In this system, we employ purified RNAPII, transcription templates and Sen1 HD and we monitor the capacity of the different versions of Sen1 to promote the release of paused RNAPIIs and associated transcripts from the DNA template (figure 2E). In agreement with the former results, the T1623V variant behaved as the wt in IVTT assays. In contrast, the T1623E mutant was dramatically affected in its capacity to dissociate the RNAPII from the DNA, both at 30°C and 37°C (figure 2E and S2). Although the *in vitro* system seems more sensitive than the *in vivo* conditions to Sen1 dysfunction, these results confirm that the T1623E phospho-mimetic mutation impairs Sen1 transcription termination activity.

Overall, the results of our *in vitro* biochemical assays provide strong support to the notion that phosphorylation of Sen1 T1623 modulates negatively the capacity of Sen1 to induce transcription termination by interfering with its association with the RNA. Furthermore, they suggest that a threonine rather than a valine at that particular position of the helicase domain might have been selected in Sen1 proteins to provide the capacity to regulate their catalytic activity by phosphorylation.

***Phosphorylation of Sen1 T1623 causes deregulation of a small subset of functionally-***

### ***related protein-coding genes***

The identification of Sen1 T1623 phosphorylation as a mechanism of regulation of Sen1-mediated transcription termination prompted us to search for the targets of such regulation. To this end, we performed genome-wide analysis of transcription in the presence of the T1623E phosphomimetic mutation, hoping that this mutation would be sufficient to reproduce specific gene expression patterns that could inform about the regulatory function of Sen1 phosphorylation. Thus, we generated high-resolution maps of transcribing RNAPII by UV crosslinking and analysis of cDNA (CRAC, Granneman et al., 2009) in yeast strains expressing either the wt or the T1623E version of *SEN1* (figure 3). The CRAC technique relies on the detection and quantification of nascent transcripts that crosslink to and co-purify with transcribing RNAPII. Experiments were performed on two independent biological replicates that yielded very reproducible results (figure S3). We first analyzed the distribution of RNAPII at non-coding genes whose transcription is normally terminated by Sen1 (i.e. snoRNA genes and CUT genes) to assess to what extent the transcription termination defects observed at individual targets in the T1623E mutant are widespread. Metagene analyses of snoRNA genes revealed a moderate increase in the RNAPII signal downstream of the annotated transcription termination site, indicating mild transcription termination defects (figure 3A). More detailed heatmap analyses of individual snoRNA genes showed that these termination defects applied to most snoRNA genes and consisted in a delay in termination of ~300 nt on average (figure 3B and representative example in figure 3C). A similar behaviour was observed at CUTs, although the termination defects were less pronounced than at snoRNA genes (figure 3D-F). Altogether, these results indicate that the T1623E mutation provokes moderate transcription termination defects at a majority of, albeit not all, non-coding genes that depend on Sen1 for termination.

We next performed a differential expression analysis on protein-coding genes, that we expected to be the ultimate targets of regulation by Sen1 T1623 phosphorylation (figure 4A). Interestingly, despite the relatively widespread nature of the termination defects observed in the *sen1T1623E* mutant, our analyses revealed significant changes in the expression of only a small (~1%) subset of protein-coding genes. Specifically, we detected 66 genes that were upregulated and 53 genes that were downregulated in the *sen1T1623E* mutant. We manually curated this gene list to exclude pseudogenes and others false positives (see methods for details) and we ended up with a more robust set of 84 genes that was subjected to more in depth analyses (table S2).

To explore the connections between the defects in non-coding transcription termination and the changes in the expression of protein-coding genes provoked by the Sen1 T1623E mutation, we performed an analysis of non-coding transcription relative to differentially-expressed (DE) protein-coding genes (figure 4B). Because genes that are regulated by non-coding transcription are typically located either downstream of or antisense to non-coding transcription units, we reasoned that genes in either or both configurations might be overrepresented in the set of DE genes compared to the whole protein-coding genome. Indeed, we observed a moderate but statistically significant enrichment in genes located just downstream of annotated non-coding genes (CUTs and Stable Unannotated Transcripts or SUTs, Z. Xu et al., 2009) whose transcription is typically terminated by the NNS-dependent pathway (Schulz et al., 2013). This suggests that at least part of the changes in gene expression we observed in the *sen1T1623E* mutant could be due to events of transcriptional interference by impaired termination of non-coding transcription just upstream of the repressed genes.

Next we performed a gene ontology analysis on the set of DE genes to search for possible signatures or functional connections between these genes. Interestingly, several common themes emerged, especially from the set of downregulated genes (figure 4C and tables S3 and S4). Indeed, GO terms related to transport, energetic metabolism and mitochondrial activity were strongly overrepresented in the list of downregulated genes. Furthermore, regarding the cellular location of the encoded proteins, the vast majority of downregulated genes corresponded to cell wall and membrane proteins and an important fraction of them were actually part of the inner mitochondrial membrane. Concerning the upregulated genes, we found a modest enrichment in genes involved in UMP synthesis, likely related to the effects of the T1623E mutation on *URA2* (figure S4), a well-characterized gene regulated by Sen1-mediated transcription termination (Thiebaut et al., 2008). We also detected an enrichment in genes encoding cell wall and mitochondrial components.

Guided by the results of our GO analysis we investigated the function of each DE gene individually and we classified genes into non-overlapping functional categories that could facilitate



subsequent analyses (figure 4D and table 1). In this way we found that roughly 40% of DE genes encoded proteins related to mitochondrial functions, about 20% of genes were assigned to cell wall and plasma membrane related functions and ~10% participated in nucleotide and amino acid biosynthetic pathways. Interestingly, we also found that ~10% of DE genes were involved in Zn homeostasis. Finally, about 20% of genes, grouped as "others", corresponded mainly to genes with unknown functions and genes encoding transcription factors or vacuolar proteins. When we inspected separately upregulated and downregulated genes, we found that genes involved in mitochondrial functions and Zn homeostasis were mostly downregulated, while the other two main functional groups were composed of more similar numbers of up- and downregulated genes.

In summary, these results suggest that phosphorylation of Sen1 T1623 could play a role in the regulation of specific pathways and cellular processes.

### ***The T1623 phospho-mimetic mutation provokes repression of the gene encoding the master regulator of Zn homeostasis by transcriptional interference***

We decided to focus on genes related to Zn homeostasis since, as mentioned above, several genes involved in this process were previously shown to be regulated by non-coding transcription (Amanda J. Bird et al., 2006). Interestingly, when we examined carefully the genes from this group, we found that the gene encoding the master regulator of Zn homeostasis, Zap1, is located just downstream of a non-annotated transcription unit (figure 5A). The transcripts produced by this putative non-coding gene behave as CUTs because RNAseq data reveal that these RNAs are strongly stabilized by deletion of *RRP6*, which encodes one of the major exonucleases responsible for degradation of CUTs (H. Neil et al., 2009; Wyers et al., 2005; Z. Xu et al., 2009). Importantly, the Sen1 T1623E phospho-mimetic mutation induced transcription termination defects at this non-coding gene, which resulted in increased transcription through the promoter region of *ZAP1* and, consequently, decreased RNAPII signal at *ZAP1* gene body and decreased levels of *ZAP1* mRNA (figure S5). These observations strongly suggest that phosphorylation of Sen1 T1623 promotes repression of the *ZAP1* gene by a mechanism of transcriptional interference mediated by impaired transcription termination of an upstream non-coding gene. We dubbed this novel non-coding gene *ZRN1* for Zap1-Repressor Non-coding gene 1 and we used our CRAC datasets together with former annotations of transcriptional start sites in *Δrrp6* (Challal et al., 2018) to define the coordinates of this gene as positions 334,108 (start) and 333,439 (end) of chromosome X.

Zap1 is a transcriptional activator that recognizes a sequence motif called Zn Responsive Elements (ZRE) at the promoter region of the genes it regulates and recruits the SWI/SNF, SAGA, and Mediator complexes to promote the expression of genes involved in Zn uptake, Zn mobilization from vacuolar stores, etc (Frey and Eide, 2012). Zap1 regulates its own expression by binding to a ZRE that is located at the region between *ZRN1* and *ZAP1* genes, suggesting that increased transcription through this region in the *sen1T1623E* mutant could impair the recognition of this sequence by Zap1 (figure 5A-B).

### ***Phosphorylation of Sen1 T1623 partially mimics conditions of Zn excess***

Repression of *ZAP1* expression could be responsible for at least part of the changes in gene expression provoked by the Sen1 T1623E mutation. Indeed, we observed a significant overlap between the set of DE genes and previously described Zap1 direct targets (Lyons et al., 2000; Wu et al., 2008). Furthermore, the expression pattern of Zap1 target genes in the *sen1T1623E* mutant is consistent with decreased levels of the Zap1 regulator. We observed, for instance, strong repression of *ZRT1*, which encodes a high-affinity Zn permease and is typically activated by Zap1, as well as repression of the non-coding gene *ZRR1* and concomitant upregulation of *ADH1*, which is under the control of *ZRR1* (figure 5D and S5). The fact that we observe a slight induction of some genes typically activated in response to Zn deficiency (e.g. *ZRT1* and *ZRR1*) indicates that in the growth conditions we have employed in our analyses (i.e. growth in minimal medium) cells experience a moderate Zn limitation, consistent with former transcriptome analyses performed in the same conditions (Toesca et al., 2011).

We next aimed at understanding to what extent the deregulation of genes involved in Zn homeostasis could be linked to the observed changes in expression in genes belonging to the other

functional categories that we have defined (e.g. mitochondrial function, cell wall and membrane related functions, etc; figure 4D). To address this question, we compared the expression pattern of our DE genes in the *sen1T1623E* mutant relative to the wt, in  $\Delta zap1$  relative to wt and in Zn excess, where *ZAP1* is naturally repressed, relative to Zn limitation using available microarray datasets (Lyons et al., 2000). An important limitation of these analyses is the different transcriptional readout provided by the datasets employed. While our CRAC data inform about the levels of nascent transcription, the microarray data provide the total transcript amount, and therefore accumulate both transcriptional and post-transcriptional levels of regulation. Notwithstanding this limitation, all the genes directly involved in Zn homeostasis presented the same trend in the three datasets, as expected (figure 5E). Strikingly, we observed that a large fraction of the remaining genes responded in a very similar way to the *Sen1 T1623E* mutation and the *ZAP1* deletion. This suggests that the main function of *Sen1* phosphorylation at T1623E could be to repress *ZAP1* expression. Notably, this observation concerned the vast majority of genes involved in mitochondrial function, which appeared mainly repressed. Many genes encoding components of the mitochondrial respiratory chain were downregulated in the two mutants as well as in wt cells under conditions of Zn excess, suggesting that under Zn excess repression of *ZAP1* expression could directly or indirectly promote a reduction of the respiratory activity (see discussion for possible explanations).

Strikingly, we observed a group of genes whose expression pattern was similar in the *sen1T1623E* and Zn-excess datasets but not in the  $\Delta zap1$  dataset. These include for instance genes involved in detoxification (e.g. *PDR16* and *PDR5*) and, most interestingly, *PHO84*. The latter is probably the best-characterized example of gene repressed by an antisense non-coding gene dependent on the NNS-pathway. Former reports showed that mutations impairing the function of the NNS-complex provoke increased transcription antisense to the *PHO84* promoter and, consequently, repression of *PHO84* expression (Castelnuovo et al., 2014, 2013). We observed the same effect in the *sen1T1623E* mutant, strongly suggesting that the phosphorylation at T1623 induces *PHO84* downregulation (figure 5F). Similar repression levels were observed under conditions of Zn excess. *PHO84* encodes a high-affinity phosphate permease, however previous studies showed that this protein can also function as a low-affinity Zn transporter (Jensen et al., 2003) and therefore, its repression in conditions of Zn excess could be necessary to minimize the accumulation of Zn, which can be toxic at high concentrations. Taken together these results suggest that the T1623E mutation partially mimics conditions of Zn excess and, thus, these conditions could be the origin of the T1623 phosphorylation (figure 6).

## DISCUSSION

Non-coding transcription plays a major role in the regulation of gene expression. In metazoans ncRNAs exist in multiple flavors and can employ a variety of transcriptional and post-transcriptional mechanisms to modulate gene expression (reviewed in Fernandes et al., 2019). However, the transcription of ncRNAs itself can also promote gene regulation, typically by a mechanism of transcriptional interference (reviewed in Pelechano and Steinmetz, 2013). Budding yeast employs almost exclusively this regulatory strategy. In most cases of repression by transcriptional interference, the non-coding gene is expressed only under the specific condition where repression of a downstream or antisense gene is required (reviewed in Colin et al., 2011). However, under all conditions studied, non-coding transcription is widespread and the genome of *S. cerevisiae* very compact, so a large number of genes are transcribed just downstream of or antisense to non-coding genes ( $\approx 22\%$  of genes, see figure 4) and could, therefore, be potentially subjected to transcriptional interference. Undesired gene repression is prevented by the NNS-complex, which induces early termination at most non-coding transcription units in yeast. Indeed, nuclear depletion of the Nrd1 component of the NNS complex was proposed to cause directly deregulation of about 300 genes by transcriptional interference as well as multiple additional indirect effects in gene expression (Gill et al., 2020; Schulz et al., 2013). This highlights the important role of the NNS-complex as a safeguard of genome expression but also raises the question whether the activity of the NNS-complex could be modulated under particular conditions to promote the regulation of specific genes.

Former studies have unveiled mechanisms that could regulate the action of the NNS-complex. For instance, it has been shown that several stress-responsive genes can be repressed by

premature termination by the NNS-complex and alternative TSS usage under stress conditions allows excluding Nrd1 and Nab3 recognition sequences from the 5' UTR and, thus, escaping NNS binding and transcription termination (Bresson et al., 2017). In addition, several phosphorylation sites have been identified in Nrd1 (Miranda M. Darby et al., 2012) and very recently several methylation sites have been detected in both Nrd1 and Nab3 (Lee et al., 2020), but a role for these modifications in the regulation of transcription termination has not been demonstrated. The levels of the Sen1 protein have shown to be modulated during the cell cycle (Mischo et al., 2018). Although the function of this modulation is not completely understood, it has been proposed that it could be important to prevent aberrant termination, possibly by adjusting the amount of termination factor to the actual transcriptional input of each phase of the cell cycle. In the present study we describe the identification and characterization of a phosphorylation at the catalytic domain of Sen1 and demonstrate that this modification modulates negatively the capacity of Sen1 to promote transcription termination. The fact that the phosphorylated residue, a threonine, is evolutionary conserved from yeast to humans together with our evidences showing that a similar but non-phosphorylatable amino acid at the same position can support Sen1 function *in vivo* and *in vitro* led us to propose that a threonine has been maintained in Sen1 proteins for regulatory purposes (figures 1-2).

Despite provoking relatively widespread effects on non-coding transcription termination, a phospho-mimetic mutation in the identified phosphosite, T1623, led to changes in the expression of only a small number of genes. Among them, very few were likely to be directly repressed by an inefficiently terminated neighboring non-coding gene (figure 4), suggesting a rather specific role for this phosphorylation in gene regulation. Indeed, our biochemical and genomic data support the idea that phosphorylation of T1623 modulates negatively the association of Sen1 with the nascent RNA, which induces a delay in transcription termination of up to few hundred base pairs. Such delay in termination can result in repression of cognate genes under particular circumstances, for instance, when there is a gene immediately downstream of the inefficiently terminated non-coding transcription unit or when the non-coding gene is terminated already sub-optimally and a mild dysfunction in the NNS-complex can provoke exceptionally strong termination defects. In such scenario, the specificity of the regulation by T1623 phosphorylation would be provided by the particular properties of the target genes. Alternatively, specificity could be achieved locally, in other words by modifying only Sen1 molecules at particular *loci*. In this respect, a growing number of kinases have been found to associate with the chromatin and to regulate transcription-related processes (Chow and Davis, 2006; Nemeč et al., 2019; Yurko et al., 2017). Sen1 could therefore be a target of one or several of these kinases. The identity of the kinase/s phosphorylating the Sen1 residue T1623 will shed light onto this question. Unfortunately, our attempt to identify the protein/s responsible for this phosphorylation using an *in vitro* screening with purified kinases was not successful (data not shown). Different approaches will be needed to accomplish this challenging goal in the future.

Among the genes that were differentially expressed in the presence of the Sen1 T1623E phospho-mimetic mutation, a group of genes involved in Zn homeostasis attracted our attention. Zn is an important protein cofactor. It has been estimated that around 10% of yeast proteins require the association with Zn for their enzymatic activity and/or structural stability (Wang et al., 2018). However, Zn excess can also be toxic for instance because it can compete with other metal ions for the active sites of enzymes. Therefore, Zn intracellular levels need to be tightly controlled, which involves mainly the regulation of Zn uptake and trafficking and the expression of Zn-associated proteins. As mentioned above, a major actor in this process is the transcriptional activator Zap1, which directly controls the expression of genes encoding Zn transporters (e.g. *ZRT1* and *ZRT2*), Zn-containing proteins as well as many genes involved in the cellular adaptation to Zn limitation (Lyons et al., 2000; Wu et al., 2008). The activity of Zap1 is highly regulated by Zn at different levels. As outlined above, Zap1 is a DNA binding protein that recognizes ZRE motifs at the promoter region of its target genes and recruits several coactivators through two different transactivation domains (Frey and Eide, 2012). At high concentrations, Zn can interact with several regions of Zap1 and modulate negatively the activity of both the transactivation domains and the DNA-binding domain (reviewed in Wilson and Bird, 2016). In addition, Zap1 is transcriptionally autoregulated, in other words, it binds to and activates its own promoter (Zhao et al., 1998). In the present study we describe an additional layer of transcriptional regulation that relies on a non-coding gene located just upstream of the *ZAP1* gene and that we have named *ZRN1* (figure 5). Unlike the other formerly reported non-coding genes involved in the regulation of Zn homeostasis, *ZRR1* and *ZRR2*, whose expression is activated in response to Zn limitation, *ZRN1* seems constitutively expressed and acquires the capacity to repress

*ZAP1* via the negative modulation of its transcription termination upon Sen1 T1623 phosphorylation. The expression pattern of Zn homeostasis genes, as well as others (figure 5 and 6) in the *sen1T1623E* phospho-mimetic mutant suggests that Sen1 would be phosphorylated under Zn excess. Both the amount of active Zap1 proteins, determined by the transcriptional output of the *ZAP1* gene and the intracellular concentration of free Zn, and the affinity of Zap1 for the different versions of its recognition motif found in the regulated promoters determines the Zn-responsiveness of the different Zap1 targets. While some genes are activated under mild Zn limitation, others require severe Zn starvation to be activated (Wu et al., 2008). The combined action of multiple transcriptional and post-transcriptional levels of regulation on Zap1 could be important to better adjust the expression levels of the different Zap1 targets to the actual cellular needs and avoid wasteful production of proteins (e.g. high-affinity Zn transporters) when they are not required (e.g. under Zn-replete/excess conditions). These different mechanisms of regulation of Zap1 activity likely act in a at least partially redundant manner. However, it possible that in a natural environment, where yeast might be subject to multiple simultaneous nutritional and physical stresses, all these mechanisms are necessary for the accurate regulation of Zn homeostasis genes which could be important for cell fitness.

In addition to the repression of *ZAP1*, the Sen1 T1623E phospho-mimetic mutation induces downregulation of the expression of *PHO84*. Repression of this gene seems independent on the effect of the Sen1 mutation on the expression of *ZAP1*, since former data did not reveal any decrease in *PHO84* expression in a  $\Delta zap1$  mutant (figure 5). However, *PHO84* seems also repressed under Zn excess (figure 5). This gene encodes a high-affinity phosphate permease that can also function as a low-affinity Zn transporter and its overexpression leads to toxicity due to heavy metal accumulation (Jensen et al., 2003; Ofiteru et al., 2012), which might explains the necessity to keep *PHO84* expression low under Zn excess. *PHO84* is one of the first and best-characterized examples of gene repressed by antisense transcription in budding yeast (Camblong et al., 2007). Transcription of the antisense non-coding gene (*CUT281*) is terminated by the NNS-complex and in different mutant contexts where the activity of the NNS-complex is compromised, impaired termination results in increased antisense transcription through the *PHO84* promoter region and, consequently, the repression of *PHO84* by transcriptional interference (Castelnuovo et al., 2014, 2013). Similarly, the phospho-mimetic T1623E mutation reduces the efficiency of transcription termination at *CUT281* and leads to an accumulation of RNAPIIs antisense to the promoter of *PHO84* as well as a decrease in *PHO84* transcription. Strikingly, the *PHO5* gene, which shares a similar configuration with an antisense NNS-dependent CUT (Castelnuovo et al., 2014) but whose function is exclusively related to the phosphate metabolism, did not exhibit the same behavior as *PHO84* (data not shown). This supports the idea that the Sen1 T1623 phosphorylation could play a specific role in the regulation of zinc homeostasis or related processes. Therefore, we propose that under Zn excess phosphorylation of Sen1 at T1623 induces the repression of *PHO84* expression to further contribute to the regulation of Zn intracellular concentration (figure 6).

The most represented category within the genes that respond to the Sen1 T1623E phospho-mimetic mutation, and, therefore, most likely to the phosphorylation of T1623, is the mitochondrial-related function. Indeed 40% of genes encoding components of the mitochondrial respiratory chain were significantly downregulated in the mutant (figure 4). Strikingly, we observed a similar behavior for many of these genes in the  $\Delta zap1$  mutant as well as in Zn excess relative to limitation (figure 5). Former data indicate that the excess of Zn can interfere with iron metabolism and provoke an increase in the production of reactive oxygen species or ROS (Pagani et al., 2007; Zhao et al., 2020). Because several components of the mitochondrial respiratory chain use iron as a cofactor (Kim et al., 2012) and mitochondria are a major source of ROS production, it is possible that an excess of Zn might affect the stability and function of mitochondrial respiratory complexes. Yeast cells might therefore adapt their energetic metabolism to conditions of Zn excess by reducing their respiratory capacity, through the repression of the components of the respiratory chain, and promoting the use of fermentative pathways for energy production. Indeed, budding yeast can use a mixed fermentative-respiratory metabolism and in the presence of high levels of glucose it employs preferentially fermentation (Rolland et al., 2002). The fact that these genes appeared also repressed in the  $\Delta zap1$  datasets, which were generated under Zn limitation, suggests that the direct and indirect effects of the lack of Zap1 protein on gene expression might be interpreted by the cell as conditions of Zn excess, where *ZAP1* is naturally repressed. In the same line, *ZAP1* repression could underlie the effect of the T1623E phospho-mimetic mutation on genes of mitochondrial components.

However, we have observed that at least one gene encoding a component of the complex V of the electron transport chain, *ATP16*, is located just downstream of an annotated non-coding gene that depends on the NNS-complex for transcription termination (CUT060, see figure S6A). A former study has shown that mutation of the Nrd1 component of the NNS-complex impairs transcription termination at CUT060 and the repression of *ATP16* expression (du Mee et al., 2018). Similarly, in the *sen1T1623E* phospho-mimetic mutant we observed a moderate increase in the RNAPII signal at the CUT060-*ATP16* intergenic region suggesting a termination defect at CUT060 that could underlie the decrease in *ATP16* transcription observed in the mutant (figure S6A). Reduced expression of the components of the respiratory chain can result in the inability of yeast cells to metabolize non-fermentable carbon sources as glycerol. Indeed, we observed that the *sen1T1623E* mutant could not grow in rich medium with glycerol as the carbon source, while a wt and a  $\Delta zap1$  mutant strains could (figure S6B). This indicates that in rich medium the lack Zap1 is not enough to induce repression of genes of mitochondrial components, or at least to reduce sufficiently the expression levels of these genes to prevent growth in glycerol. Moreover, this results strongly suggest that the Sen1 T1623E phospho-mimetic mutation promotes the repression of *ATP16* expression independently on the regulation of *ZAP1* expression.

This data supports a model according to which under conditions of Zn excess, phosphorylation of Sen1 T1623 would promote the repression of *ZAP1* and *PHO84*, to regulate the intracellular concentration of Zn, as well as the downregulation of components of the mitochondrial respiratory chain to contribute to the appropriate metabolic adaptations.

Overall, our results expand the functional dimension of pervasive transcription by showing that modulation of the efficiency of non-coding transcription termination can promote the regulation of gene expression.

## METHODS

### Protein sequence alignments

The sequence of Sen1 protein from *Saccharomyces cerevisiae* (Q00416) was aligned with that of the orthologues from *Candida albicans* (XP\_723150), *Schizosaccharomyces pombe* (Q92355), *Mus musculus* (A2AKX3), zebrafish (E7FBJ2), and *Homo sapiens* (Q7Z333) as well as with Upf1 sequence from *S. cerevisiae* (P30771) using clustal omega (<https://www.ebi.ac.uk/Tools/msa/clustalo>). Sequence alignment was visualized with the jalview software.

### Protein immunoprecipitation experiments and mass spectrometry analyses

Yeast extracts were prepared by standard methods. Briefly, cell pellets were resuspended in lysis buffer (10 mM sodium phosphate pH 7, 200 mM sodium acetate, 0.25% NP-40, 2 mM EDTA, 1 mM EGTA, 5% glycerol) containing protease inhibitors, frozen in liquid nitrogen and lysed using a Retsch MM301 Ball Mill. Protein extracts were clarified by centrifugation at 13000 rpm for 45 min at 4°C and incubated with RNase A (20 µg/ml) for 20 min at RT. Extracts were then incubated with IgG-coupled magnetic beads (Dynabeads, Thermo Fisher, 30 mg/ml) for 2h at 4°C and beads were extensively washed with lysis buffer. Beads were finally washed with H<sub>2</sub>O and bound proteins were directly subjected to trypsin digestion and mass spectrometry analysis.

### Construction of yeast strains and plasmids

Yeast strains used in this paper are listed in table S5. Gene tagging was performed with standard procedures (Longtine et al., 1998; Rigaut et al., 1999) using plasmids described in table S6. Strains DLY3381 and DLY3340 expressing *sen1T1623V* and *sen1T1623E*, respectively were constructed by transforming a  $\Delta sen1$  strain harbouring the *URA3*-containing plasmid pFL38-*SEN1* (DLY2767) with the product of cleavage of pFL39-*sen1T1623V* (pDL970) or pFL39-*sen1T1623E* (pDL921) with restriction enzymes MluI, BstZ17I and Bsu36I. Cells capable of growing on 5-FOA were then screened by the absence of pFL38-*SEN1* and pFL39-derived plasmids and the presence

of the desired mutation in the chromosome was verified by DNA sequencing.

Plasmids used in this study are listed in table S6. Plasmids expressing different *SEN1* variants were constructed by homologous recombination in yeast. Briefly, a DNA fragment harbouring the desired mutation was obtained by mutagenic overlapping PCR. Then, a wt yeast strain was transformed with a linearized vector and the corresponding PCR fragment, which contains 40–45 bp sequences at both ends allowing recombination with the vector. Clones were screened by PCR, and the positive ones were verified by sequencing.

Plasmids for overexpression of the T1623V and T1623E variants of Sen1 HD in *E. coli* were generated by introducing the corresponding mutations in plasmid pDL893 using the QuikChange site-directed mutagenesis kit (Stratagene).

### Northern blot assays

Yeast cells used for northern blot assays were grown on the appropriate medium, depending on the experiment, to OD<sub>600</sub> 0.3 to 0.6. Unless otherwise indicated, cells were grown at 30°C. Cells were harvested by centrifugation and RNAs were prepared using standard methods. Samples were separated by electrophoresis on 1.2% agarose gels, and then transferred to nitrocellulose membranes and UV-crosslinked. Radiolabeled probes were typically generated by random priming of PCR products covering the regions of interest with Megaprime kit (GE Healthcare) in the presence of  $\alpha$ -<sup>32</sup>P dCTP (3000 Ci/mmol). Oligonucleotides used to generate the PCR probes are indicated in table S7. Hybridizations were performed using a commercial buffer (UltraHyb, Ambion) at 42°C and after extensive washes, membranes were analysed by phosphorimaging.

### Protein expression and purification

The different Sen1 HD versions were fused to a C-terminal His<sub>8</sub>-tag coupled to *Vibrio cholerae* MARTX toxin cysteine protease domain (Shen et al., 2009) and expressed from the T7 promoter (see plasmids in table S6) in *Escherichia coli* strain BL21 (DE3) CodonPlus (Stratagene). Overexpression was induced by growth in auto-inducing medium (Studier, 2005) at 20°C overnight. Protein extracts were prepared by sonication of the resuspended cell pellets and subsequent clarification by centrifugation at 20,000 *g* for 30 min at 4°C. Cells were lysed in binding buffer containing 50 mM Tris-HCl pH 8, 500 mM NaCl, 2 mM MgCl<sub>2</sub>, 10 mM imidazole, 10% (v/v) glycerol, 1 mM DTT, benzonase, and protease inhibitors. The proteins were bound to a Ni-NTA sepharose beads (Qiagen), extensively washed with binding buffer and eluted in buffer containing 50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM MgCl<sub>2</sub>, 5% glycerol and 1 mM DTT. The eluates were then subjected to heparin affinity chromatography on a HiTrap Heparin HP column (GE Healthcare) using buffer A for binding (50 mM Tris-HCl pH 7.5, 200 mM NaCl, 2 mM MgCl<sub>2</sub>, 5% glycerol and 1 mM DTT) and buffer B for elution (50 mM Tris-HCl pH 7.5, 1 M NaCl, 2 mM MgCl<sub>2</sub>, 5% glycerol and 1 mM DTT). Proteins were dialysed against buffer containing 50 mM Tris-HCl, pH 7.5, 350 mM NaCl, 2 mM MgCl<sub>2</sub>, 50% glycerol and 1 mM DTT and stored at -80°C.

RNAPII core enzyme was purified from *S. cerevisiae* strain BJ5464 (Kireeva et al., 2003) by Ni<sup>2+</sup>-affinity chromatography followed by anion exchange on a MonoQ column basically as previously described (Porrúa and Libri, 2015c). Recombinant His<sub>6</sub>-tagged Rpb4/7 heterodimer was purified from *E. coli* by Ni<sup>2+</sup>-affinity chromatography followed by gel filtration as previously described (Porrúa and Libri, 2015c).

### Electrophoretic mobility shift assays

Experiments were performed in 10  $\mu$ l-reactions containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 2 mM MgCl<sub>2</sub>, 7.5  $\mu$ M ZnCl<sub>2</sub>, 10% glycerol and 1 mM DTT. The RNA substrate (4 nM final concentration) was incubated with increasing concentrations of Sen1 HD variants at either 30°C or 37°C for 15 min. Reactions were loaded onto 5% native polyacrylamide gels and the different RNA species were separated by electrophoresis in 0.5xTBE at 100V for 1h. Gels were scanned directly using a Typhoon scanner (GE Healthcare) and analyzed with the ImageQuant TL software. Data were fit

using a non-linear regression model and the  $K_d$  (dissociation constant) for each protein and condition was calculated using the Prism software (version 9.0.0).

### **ATP hydrolysis assays**

ATPase assays were performed essentially as previously described (Porrúa and Libri, 2015c). Briefly, 10 nM of purified Sen1 proteins was assayed at 30°C or 37°C in 10- $\mu$ l reactions containing 10 mM Tris–HCl pH 7.5, 75 mM NaCl, 1 mM MgCl<sub>2</sub>, 1 mM DTT, 15% glycerol, and 0.5 ng/ $\mu$ l polyU. The reaction started with the addition of a 250  $\mu$ M ATP solution containing 0.25  $\mu$ M of 3000 Ci/mmol  $\alpha$ <sup>32</sup>P-ATP as the final concentrations. Aliquots were taken at different time points, mixed with 2.5 volumes of quench buffer containing 10 mM EDTA and 0.5% SDS, and then subjected to thin-layer chromatography on PEI cellulose plates (Merck) in 0.35 M potassium phosphate (pH 7.5). The different radioactive species were analyzed by phosphorimaging.

### ***In vitro* transcription termination assays**

Termination assays were performed basically as previously described (Porrúa and Libri, 2015c). Briefly, ternary ECs were assembled in a promoter-independent manner by first annealing a fluorescently labeled RNA (oligo DL2492, see table S7) with the template DNA (oligo DL3352, see table S7) and subsequently incubating the RNA:DNA hybrid with purified RNAPII. Next, the non-template (NT) strand (oligo DL3353, see table S7) and recombinant Rpb4/7 heterodimer were sequentially added to the mixture. The ternary ECs were then immobilized on streptavidin beads (Dyna-beads MyOne Streptavidin T1 from Invitrogen) and washed with transcription buffer (TB) containing 20 mM Tris–HCl pH 7.5, 100 mM NaCl, 8 mM MgCl<sub>2</sub>, 10  $\mu$ M ZnCl<sub>2</sub>, 10% glycerol, and 2 mM DTT; then with TB containing 0.1% Triton, TB containing 0.5 M NaCl, and finally TB. The termination reactions were performed at 30°C or 37°C in TB in a final volume of 20  $\mu$ l in the absence or in the presence of 50 nM of different versions of Sen1 HD. Transcription was initiated after addition of a mixture of ATP, UTP, and CTP (1 mM each as the final concentration in the reaction) to allow transcription through the G-less cassette up to the first G of a G-stretch in the NT strand. The reactions were allowed for 15 min and then stopped by the addition of 1  $\mu$ l of 0.5 M EDTA. After separation of beads and supernatant fractions, beads fractions were resuspended in 8  $\mu$ l of loading buffer (1 $\times$  Tris-borate-EDTA, 8 M urea) and boiled for 5 min at 95°C. Transcripts in the supernatant fractions were ethanol precipitated and resuspended in 8  $\mu$ l of loading buffer. Transcripts were separated by 10% (w/v) denaturing PAGE (8 M urea), and gels were scanned with a Typhoon scanner (GE Healthcare) and analyzed with the ImageQuant TL software (GE Healthcare).

### **UV crosslinking and analysis of cDNA (CRAC)**

For RNAPII CRAC experiments we employed the procedure reported in Granneman *et al* (Granneman *et al.*, 2009) with several modifications described in Candelli *et al* (Candelli *et al.*, 2018a). We used strains expressing an His6-TEV-Protein A-tagged version of the Rpb1 subunit of RNAPII at the endogenous locus and either the wt or the T1623E version of Sen1 (see table S5). Briefly, 2 l of yeast cells were grown at 30°C to OD<sub>600</sub> ~ 0.6 in CSM-TRP medium and crosslinked for 50 s using a W5 UV crosslinking unit (UVO3 Ltd). Cells were then harvested by centrifugation at 12,000 *g* for 10 min and the pellets were washed with cold 1 $\times$  PBS and resuspended in 2.4 ml of TN150 buffer (50 mM Tris pH 7.8, 150 mM NaCl, 0.1% NP-40 and 5 mM  $\beta$ -mercaptoethanol) containing protease inhibitors (Complete EDTA-free Protease Inhibitor Cocktail) per gram of cells. Suspensions were flash-frozen in droplets and cells were lysed using a Ball Mill MM 400 (five cycles of 3 min at 20 Hz). The mixtures were incubated with 165 units of DNase I (NEB) for 1 h at 25°C and then clarified by centrifugation at 20,000 *g* for 20 min at 4°C.

Protein extracts were subjected to IgG affinity purification on M-280 tosylactivated dynabeads coupled with rabbit IgGs (15  $\mu$ g of beads per sample). After extensive washes with TN1000 buffer (50 mM Tris pH 7.8, 1 M NaCl, 0.1% NP-40 and 5 mM  $\beta$ -mercaptoethanol), the protein–RNA complexes were eluted by digestion with the TEV protease and treated with 0.2 U of RNase cocktail (RNase-IT, Agilent) to reduce the size of the nascent RNA (note that the 3' end of nascent transcripts is protected from degradation). The eluates were mixed with guanidine–HCl to a final concentration

of 6 M and incubated with Ni-NTA sepharose (Qiagen, 100 µl of slurry per sample) o/n at 4°C. After washing beads, sequencing adaptors were ligated to the RNA molecules as described in the original procedure. Protein–RNA complexes were eluted with 400 µl of elution buffer (50 mM Tris pH 7.8, 50 mM NaCl, 150 mM imidazole, 0.1% NP-40, 5 mM β-mercaptoethanol) and concentrated using Vivacon ultrafiltration spin columns. Then, proteins were fractionated using a Gel Elution Liquid Fraction Entrapment Electrophoresis (GelFree) system (Expedeon) following manufacturer's specifications and the different fractions were monitored for the presence of Rpb1 by SDS–PAGE. The fractions of interest were treated with 100 µg of proteinase K, and RNAs were purified and reverse-transcribed using reverse transcriptase Superscript IV (Invitrogen).

The cDNAs were amplified by PCR using LA Taq polymerase (Takara), and then, the PCR reactions were treated with 200 U/ml of Exonuclease I (NEB) for 1 h at 37°C. Finally, the DNA was purified using NucleoSpin columns (Macherey-Nagel) and sequenced on a NextSeq 500 Illumina sequencer.

### Data processing

CRAC reads were demultiplexed using the pyBarcodeFilter script from the pyCRACutility suite (Webb et al., 2014). The 5' adaptor was clipped with Cutadapt and the resulting insert was quality-trimmed from the 3' end using Trimmomatic rolling mean clipping (Bolger et al., 2014). We used the pyCRAC script pyFastqDuplicateRemover to collapse PCR duplicates using a six-nucleotide random tag included in the 3' adaptor. The resulting sequences were reverse complemented with the Fastx reverse complement that is part of the fastx toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)) and mapped to the R64 genome with bowtie2 using “-N 1” option. Coverage files were normalized to 10<sup>7</sup> counts.

### Bioinformatic and statistical analyses

Bioinformatic analyses were mainly performed using the Galaxy (<http://galaxy.sb-roscoff.fr>) and the R frameworks. For metagene analyses, we used the deepTools2 package (Ramírez et al., 2016). Strand-specific coverage bigwig files were used as inputs for the computeMatrix tool together with separate annotations for each strand and for each feature (CUTs or snoRNAs), using a bin size of 10 and the TTS as the reference point. Matrices generated for each strand were subsequently combined using the rbind option of the computeMatrixOperations tool and used as the input for the plotProfile tool. We typically represented the median instead of the mean values to minimize the bias towards the very highly expressed features. The log<sub>2</sub> FC of the RNAPII signal in the *sen1T1623E* mutant relative to the wt was calculated using the tool bigwigCompare. As for metagene analyses, we generated separate matrices for each strand with the log<sub>2</sub> FC bigwig files and the computeMatrix tool, and we subsequently combined them using the computeMatrixOperations tool. Heatmaps were obtained using these matrices as the input for the plotHeatmap tool. For metagene and heatmap analyses we employed the refined CUTs and snoRNA annotations reported in Han et al (Han et al., 2020b).

Differential expression analyses were performed by counting the reads from two independent biological replicates that mapped at protein-coding genes using the htseq-count tool (Anders et al., 2015) with the union mode. The count files thus generated were used as input files for the SARTools DESeq2 package (Varet et al., 2016) with default parameters.

For the analysis of non-coding transcription relative to protein-coding genes we used the CUTs and SUTs annotations reported in Xu et al, (Z. Xu et al., 2009) and the mRNAs annotations in Challal et al (Challal et al., 2018). We employed the window bed tool of the bedtools package (Quinlan and Hall, 2010) to detect the genes that overlap with an antisense CUT or SUT or that initiate within a 500-bp window downstream of one of these non-coding genes. When a gene was located both antisense to and downstream of a CUT or SUT we assigned a value of 0.5 gene to each category. To calculate the statistical significance of the enrichment of one particular category of genes (e.g. genes located just downstream of annotated non-coding genes) we performed a Fisher's exact test with R.

Gene ontology analyses were performed with the Genome Ontology Term Finder tool of the



Saccharomyces Genome Database (<https://www.yeastgenome.org/goTermFinder>).

Heatmap analyses of the expression pattern of DE genes in different backgrounds and conditions were performed with the pheatmap function of R. As input files we used the results of our differential expression analyses together with microarray datasets reported in (Lyons et al., 2000).

## ACKNOWLEDGEMENTS

We thank G. Wentzinger for technical assistance and B. Palancade, D. Libri and other members of the Libri lab for fruitful discussions. We thank D. Eide for sharing the *Δzap1* strain. We thank the Roscoff Bioinformatics platform ABiMS (<http://abims.sb-roscoff.fr>) for providing computational resources and support. This work has benefited from the facilities and expertise of the high throughput sequencing core facility of I2BC (<http://www.i2bc.paris-saclay.fr/>) as well as the expertise of the proteomic facility of the IJM. We thank C. Sole and F. Posas for their attempt to identify the kinases phosphorylating Sen1 helicase domain.

O.P was supported by the CNRS and the Agence National pour la Recherche (ANR-16-CE12-0001-01) and N.H was supported by the ANR.

## REFERENCES

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Arndt, K.M., and Reines, D. (2015). Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annual Review of Biochemistry* 84, 381–404.
- Bird, A.J., Gordon, M., Eide, D.J., and Winge, D.R. (2006). Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. *EMBO J* 25, 5726–5734.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bresson, S., Tuck, A., Staneva, D., and Tollervey, D. (2017). Nuclear RNA Decay Pathways Aid Rapid Remodeling of Gene Expression in Yeast. *Mol. Cell* 65, 787–800.e5.
- Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. (2007). Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* 131, 706–717.
- Candelli, T., Gros, J., and Libri, D. (2018a). Pervasive transcription fine-tunes replication origin activity. *Elife* 7.
- Candelli, T., Challal, D., Briand, J.-B., Boulay, J., Porrua, O., Colin, J., and Libri, D. (2018b). High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. *EMBO J*. 37.
- Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F., and Zenklusen, D. (2013). Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat. Struct. Mol. Biol.* 20, 851–858.
- Castelnuovo, M., Zaugg, J.B., Guffanti, E., Maffioletti, A., Camblong, J., Xu, Z., Clauder-Münster, S., Steinmetz, L.M., Luscombe, N.M., and Stutz, F. (2014). Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res.* 42, 4348–4362.
- Challal, D., Barucco, M., Kubik, S., Feuerbach, F., Candelli, T., Geoffroy, H., Benaksas, C., Shore, D., and Libri, D. (2018). General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. *Mol. Cell* 72, 955–969.e7.
- Chow, C.-W., and Davis, R.J. (2006). Proteins kinases: chromatin-associated enzymes? *Cell* 127, 887–890.
- Colin, J., Libri, D., and Porrua, O. (2011). Cryptic transcription and early termination in the control of gene expression. *Genet Res Int* 2011, 653494.
- Darby, M.M., Serebreni, L., Pan, X., Boeke, J.D., and Corden, J.L. (2012). The *Saccharomyces cerevisiae* Nrd1-Nab3 transcription termination pathway acts in opposition to Ras signaling and mediates response to nutrient depletion. *Mol. Cell. Biol.* 32, 1762–1775.
- Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21, 3433–3434.
- Fernandes, J.C.R., Acuña, S.M., Aoki, J.I., Floeter-Winter, L.M., and Muxel, S.M. (2019). Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Noncoding RNA* 5.

Finkel, J.S., Chinchilla, K., Ursic, D., and Culbertson, M.R. (2010). Sen1p Performs Two Genetically Separable Functions in Transcription and Processing of U5 Small Nuclear RNA in *Saccharomyces cerevisiae*. *Genetics* 184, 107–118.

Frey, A.G., and Eide, D.J. (2012). Zinc-responsive coactivator recruitment by the yeast Zap1 transcription factor. *Microbiologyopen* 1, 105–114.

Gill, J.K., Maffioletti, A., García-Molinero, V., Stutz, F., and Soudet, J. (2020). Fine Chromatin-Driven Mechanism of Transcription Interference by Antisense Noncoding Transcription. *Cell Rep* 31, 107612.

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

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.

Han, Z., Jasnovidova, O., Haidara, N., Tudek, A., Kubicek, K., Libri, D., Stefl, R., and Porrua, O. (2020). Termination of non-coding transcription in yeast relies on both an RNA Pol II CTD interaction domain and a CTD-mimicking region in Sen1. *EMBO J.* e101548.

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., and Stefl, R. (2011). Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J. Biol. Chem.* 286, 3645–3657.

Huber, F., Bunina, D., Gupta, I., Khmelinskii, A., Meurer, M., Theer, P., Steinmetz, L.M., and Knop, M. (2016). Protein Abundance Control by Non-coding Antisense Transcription. *Cell Rep* 15, 2625–2636.

Jensen, L.T., Ajuja-Alemanji, M., and Culotta, V.C. (2003). The *Saccharomyces cerevisiae* high affinity phosphate transporter encoded by PHO84 also functions in manganese homeostasis. *J Biol Chem* 278, 42036–42040.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with pervasive transcription. *Mol. Cell* 52, 473–484.

Kim, H.J., Khalimonchuk, O., Smith, P.M., and Winge, D.R. (2012). Structure, function, and assembly of heme centers in mitochondrial respiratory complexes. *Biochim Biophys Acta* 1823, 1604–1616.

Kireeva, M.L., Lubkowska, L., Komissarova, N., and Kashlev, M. (2003). Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from *Saccharomyces cerevisiae*. *Meth. Enzymol.* 370, 138–155.

Lee, K.Y., Chopra, A., Burke, G.L., Chen, Z., Greenblatt, J.F., Biggar, K.K., and Meneghini, M.D. (2020). A crucial RNA-binding lysine residue in the Nab3 RRM domain undergoes SET1 and SET3-responsive methylation. *Nucleic Acids Res* 48, 2897–2911.

Leonaitė, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O., and Conti, E. (2017). Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. *EMBO J.* 36, 1590–1604.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* 14, 953–961.

- Lunde, B.M., Hörner, M., and Meinhart, A. (2011). Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic Acids Res.* 39, 337–346.
- Lyons, T.J., Gasch, A.P., Gaither, L.A., Botstein, D., Brown, P.O., and Eide, D.J. (2000). Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast. *Proc Natl Acad Sci U S A* 97, 7957–7962.
- MacGilvray, M.E., Shishkova, E., Place, M., Wagner, E.R., Coon, J.J., and Gasch, A.P. (2020). Phosphoproteome Response to Dithiothreitol Reveals Unique Versus Shared Features of *Saccharomyces cerevisiae* Stress Responses. *J Proteome Res* 19, 3405–3417.
- Martens, J.A., Wu, P.-Y.J., and Winston, F. (2005). Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes Dev.* 19, 2695–2704.
- du Mee, D.J.M., Ivanov, M., Parker, J.P., Buratowski, S., and Marquardt, S. Efficient termination of nuclear lncRNA transcription promotes mitochondrial genome maintenance. *ELife* 7.
- 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. *Molecular Cell* 70, 312–326.e7.
- Nadal-Ribelles, M., Solé, C., Xu, Z., Steinmetz, L.M., de Nadal, E., and Posas, F. (2014). Control of Cdc28 CDK1 by a stress-induced lncRNA. *Mol Cell* 53, 549–561.
- Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038–1042.
- Nemec, C.M., Singh, A.K., Ali, A., Tseng, S.C., Syal, K., Ringelberg, K.J., Ho, Y.-H., Hintermair, C., Ahmad, M.F., Kar, R.K., et al. (2019). Noncanonical CTD kinases regulate RNA polymerase II in a gene-class-specific manner. *Nat Chem Biol* 15, 123–131.
- Nevers, A., Doyen, A., Malabat, C., Néron, B., Kergrohen, T., Jacquier, A., and Badis, G. (2018). Antisense transcriptional interference mediates condition-specific gene repression in budding yeast. *Nucleic Acids Res.* 46, 6009–6025.
- Ofiteru, A.M., Ruta, L.L., Rotaru, C., Dumitru, I., Ene, C.D., Neagoe, A., and Farcasanu, I.C. (2012). Overexpression of the PHO84 gene causes heavy metal accumulation and induces Ire1p-dependent unfolded protein response in *Saccharomyces cerevisiae* cells. *Appl Microbiol Biotechnol* 94, 425–435.
- Pagani, M.A., Casamayor, A., Serrano, R., Atrian, S., and Ariño, J. (2007). Disruption of iron homeostasis in *Saccharomyces cerevisiae* by high zinc levels: a genome-wide study. *Mol Microbiol* 65, 521–537.
- Pelechano, V., and Steinmetz, L.M. (2013). Gene regulation by antisense transcription. *Nat. Rev. Genet.* 14, 880–893.
- Porrua, 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.
- Porrua, O., and Libri, D. (2015a). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* 16, 190–202.
- Porrua, O., and Libri, D. (2015b). Characterization of the mechanisms of transcription termination by the helicase Sen1. *Methods Mol. Biol.* 1259, 313–331.

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

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851–1854.

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

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160-165.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032.

Rolland, F., Winderickx, J., and Thevelein, J.M. (2002). Glucose-sensing and -signalling mechanisms in yeast. *FEMS Yeast Research* 2, 183–201.

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*.  
Shen, A., Lupardus, P.J., Morell, M., Ponder, E.L., Sadaghiani, A.M., Garcia, K.C., and Bogoy, M. (2009). Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. *PLoS One* 4, e8119.

Soudet, J., Gill, J.K., and Stutz, F. (2018). Noncoding transcription influences the replication initiation program through chromatin regulation. *Genome Res* 28, 1882–1893.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234.

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

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

Toesca, I., Nery, C.R., Fernandez, C.F., Sayani, S., and Chanfreau, G.F. (2011). Cryptic transcription mediates repression of subtelomeric metal homeostasis genes. *PLoS Genet* 7, e1002163.

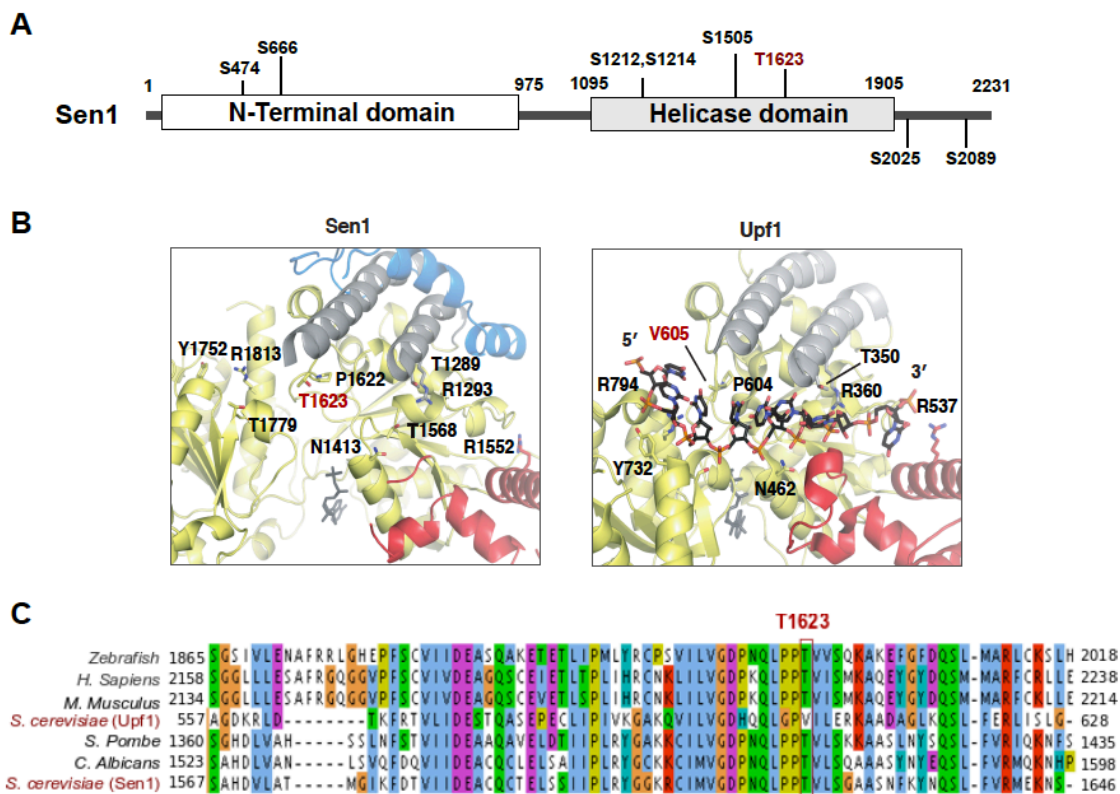
Ursic, D., Chinchilla, K., Finkel, J.S., and Culbertson, M.R. (2004). Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res.* 32, 2441–2452.

Varet, H., Brillet-Guéguen, L., Coppée, J.-Y., and Dillies, M.-A. (2016). SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS ONE* 11, e0157022.

Wang, Y., Weisenhorn, E., MacDiarmid, C.W., Andreini, C., Bucci, M., Taggart, J., Banci, L., Russell, J., Coon, J.J., and Eide, D.J. (2018). The cellular economy of the *Saccharomyces cerevisiae* zinc proteome. *Metallomics* 10, 1755–1776.

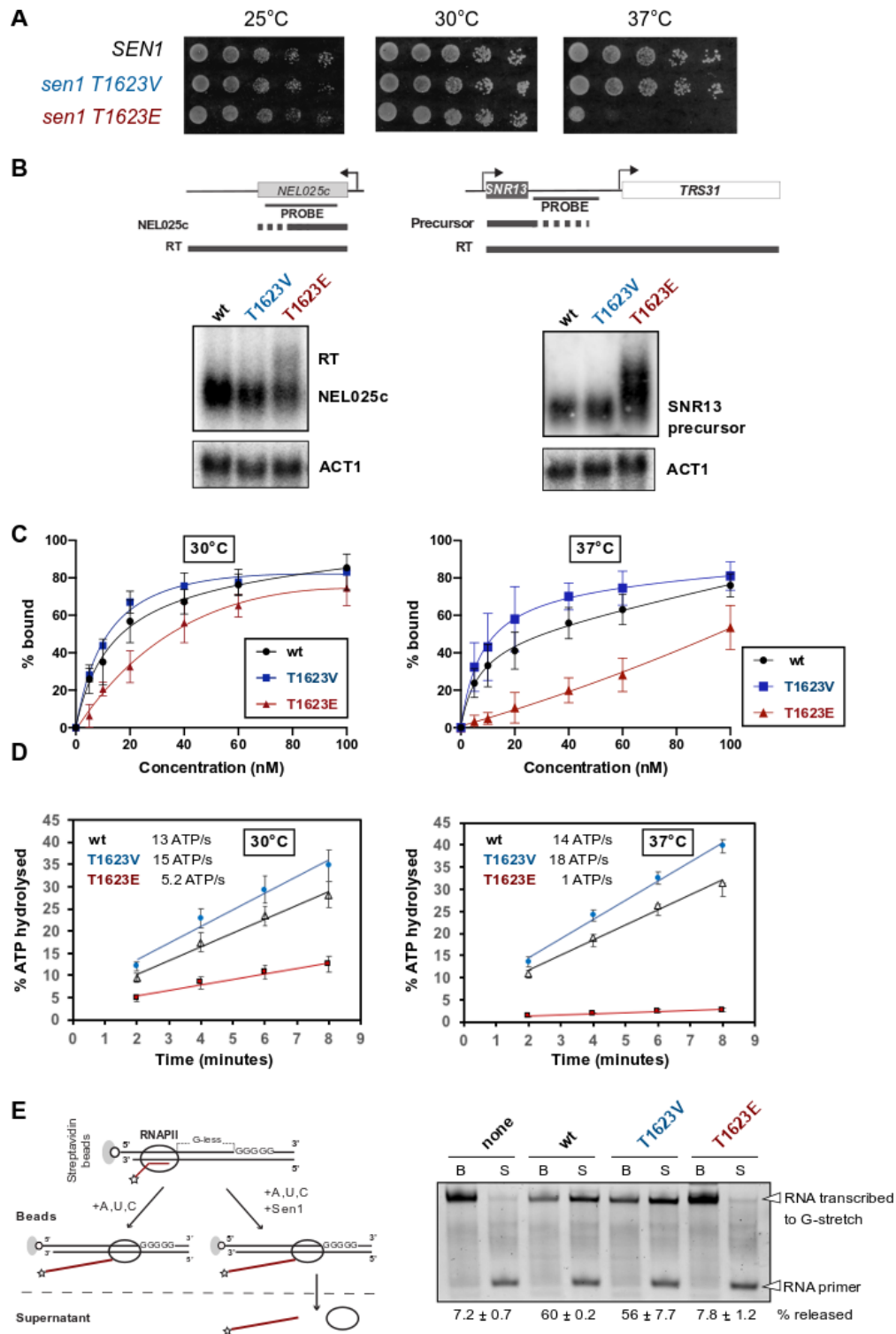
- Webb, S., Hector, R.D., Kudla, G., and Granneman, S. (2014). PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* 15, R8.
- van Werven, F.J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., and Amon, A. (2012). Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* 150, 1170–1181.
- Wilson, S., and Bird, A.J. (2016). Zinc sensing and regulation in yeast model systems. *Arch Biochem Biophys* 611, 30–36.
- Wlotzka, W., Kudla, G., Granneman, S., and Tollervey, D. (2011). The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J.* 30, 1790–1803.
- Wu, C.-Y., Bird, A.J., Chung, L.M., Newton, M.A., Winge, D.R., and Eide, D.J. (2008). Differential control of Zap1-regulated genes in response to zinc deficiency in *Saccharomyces cerevisiae*. *BMC Genomics* 9, 370.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 725–737.
- 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.
- Yurko, N., Liu, X., Yamazaki, T., Hoque, M., Tian, B., and Manley, J.L. (2017). MPK1/SLT2 Links Multiple Stress Responses with Gene Expression in Budding Yeast by Phosphorylating Tyr1 of the RNAP II CTD. *Molecular Cell* 68, 913-925.e3.
- Zhao, H., Butler, E., Rodgers, J., Spizzo, T., Dueterhoeft, S., and Eide, D. (1998). Regulation of zinc homeostasis in yeast by binding of the ZAP1 transcriptional activator to zinc-responsive promoter elements. *J Biol Chem* 273, 28713–28720.
- Zhao, Y.-Y., Cao, C.-L., Liu, Y.-L., Wang, J., Li, J., Li, S.-Y., and Deng, Y. (2020). Identification of the Genetic Requirements for Zinc Tolerance and Toxicity in *Saccharomyces cerevisiae*. *G3 (Bethesda)* 10, 479–488.

## FIGURES



**Figure 1: The catalytic domain of Sen1 is phosphorylated at a conserved threonine at the predicted RNA-binding region.**

**A)** Scheme of Sen1 protein indicating the location of the different phosphosites identified by mass spectrometry. Globular domains are indicated by solid bars whereas regions predicted to be disordered by IUPred (Dosztányi et al., 2005) are shown by a line. Phosphorylation of residue S1505 was also previously detected in former studies (MacGilvray et al., 2020; Danielle L. Swaney et al., 2013). **B)** Zoom in view of the RNA-binding region of Sen1 and Upf1. Protein regions are colored according to the color scheme used for the different subdomains of Sen1 and Upf1 helicase domains in a former study (Leonaitė et al., 2017). Specifically, amino acids in the core RecA subdomains are indicated in yellow, whereas amino acids in the so-called brace, stalk and prong are shown in blue, grey and red, respectively. Structural data correspond to PDB: 5MZN for Sen1 and PDB: 2XZO for Upf1-RNA complex. Sen1 T1623 and the equivalent V605 in Upf1 are indicated in red. Residues labeled in black are conserved positions involved in RNA binding by Upf1 that revealed a role in Sen1 interaction with RNA in a former mutational analysis (Leonaitė et al., 2017). **C)** Alignment of a small fragment of Sen1 and Upf1 proteins containing T1623 or the equivalent residues. Amino acids are colored according to the clustal color scheme.

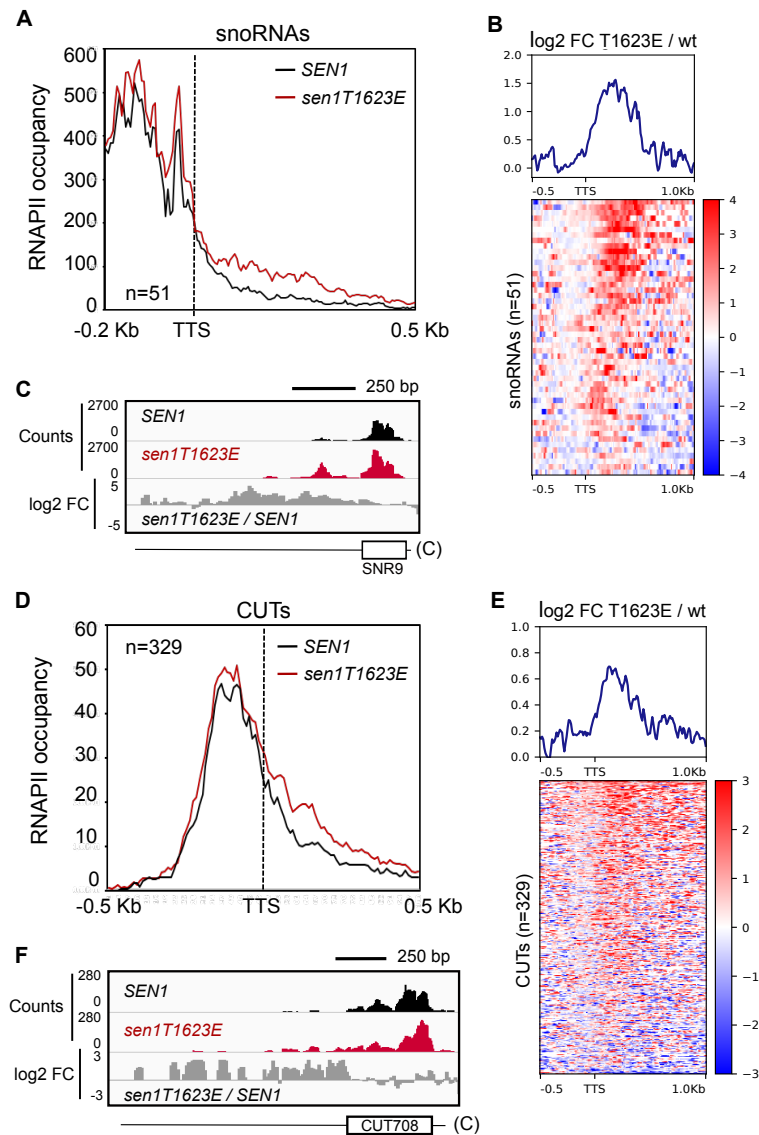


**Figure 2: Functional analyses of mutations preventing (T to V) or mimicking (T to E) phosphorylation of Sen1 T1623.**

**A)** Growth tests of yeast cells expressing different *SEN1* variants on YPD medium at the indicated temperatures. **B)** Northern blot analysis of two typical *Sen1* targets in strains expressing either the wt, the T1623V or the T1623E version of *SEN1*. Cells were grown at 30°C. Experiments were performed in a  $\Delta rrp6$  background to stabilize and, thus, render detectable the primary products of termination by *Sen1*. The *ACT1* mRNA is used as a loading control. Probes used for RNA detection are described in table S7. **C)** Electrophoretic mobility shift assay (EMSA) using a 5'-end fluorescently labeled 40-mer RNA as the substrate (DL3508, see table S7) at 4 nM and recombinant *Sen1* HD variants at 5, 10, 20, 40, 60, and 100 nM at the final concentrations. The plots correspond to the quantifications of three independent experiments performed for each incubation temperature. The values indicate the mean and standard deviation (SD). The dissociation constant (Kd) at 30°C is 13.1 nM for the wt, 12.7 nM for the T1623V and 100 nM for the T1623E version of *Sen1* HD. At 37°C



the Kds are 6.4 nM, 8.1 nM and  $> 5\mu\text{M}$  for the wt, T1623V and T1623E, respectively. Representative gels of these experiments are shown in figure S2B. **D)** Time course analysis of the nucleic acid-dependent ATPase activity of the indicated Sen1 HD variants at either 30°C or 37°C. Values represent the average and SD from three independent experiments. **E)** *In vitro* transcription termination (IVTT) assays with the indicated versions of Sen1 HD Left: Scheme of an IVTT assay. Ternary elongation complexes (ECs) composed of purified RNAPII, fluorescently labeled nascent RNA, and DNA templates are assembled and attached to streptavidin beads via the 5' biotin of the non-template strand to allow subsequent separation of beads-associated (B) and supernatant (S) fractions. A star denotes the presence of a FAM at the 5' end of the RNA. The transcription template contains a G-less cassette followed by a G-stretch in the non-template strand. After addition of a mixture of ATP, UTP and CTP at a final concentration of 1 mM each, RNAPII transcribes until it encounters the G- stretch. Sen1 HD dissociates ECs stalled at the G-stretch, therefore releasing RNAPIIs and the associated transcripts to the supernatant. The fraction of transcripts released from ECs stalled at the G-stretch is used as a measure of the efficiency of termination. Right: Representative gel of one out of two independent IVTT assays. The values on the bottom correspond to the average and the SD of three independent experiments.

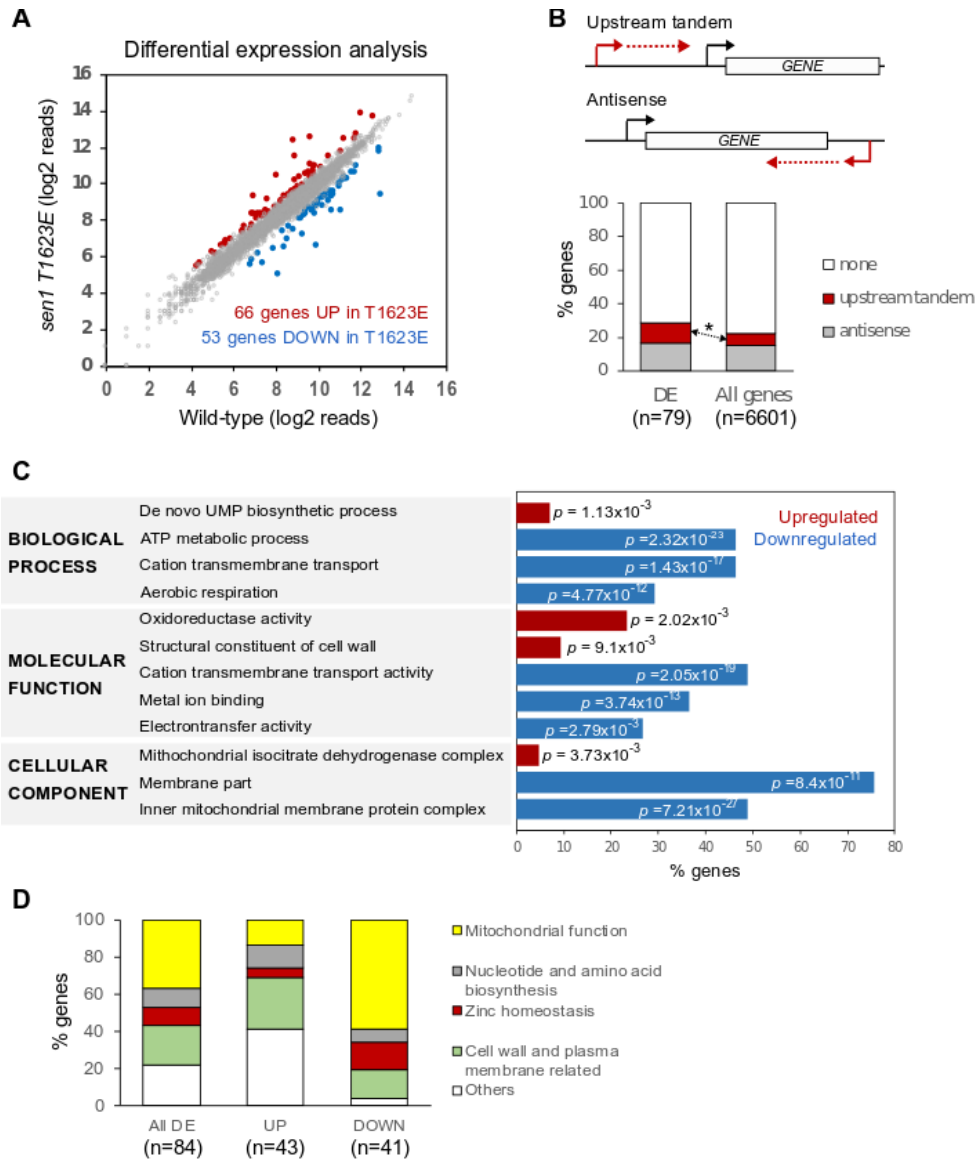


**Figure 3: A phospho-mimetic mutation in Sen1 T1623 globally reduces the efficiency of transcription termination at non-coding genes**

**A)** Metagenesis of the RNAPII distribution relative to the annotated transcription termination site (TTS) of snoRNAs. Values on the y-axis correspond to the median coverage.

**B)** Heatmap analysis representing the log<sub>2</sub> of the fold change (FC) of the RNAPII signal in the *sen1T1623E* mutant relative to the wt at snoRNA genes. The summary plot on the top was calculated using the average values for each position. **C)** Integrative Genomics Viewer (IGV) screenshots of an example of snoRNA genes displaying termination defects in the *sen1T1623E* mutant. The values correspond to the counts per million of counts (CPM) multiplied by 10. **D)** Metagenesis analyses of transcribing RNAPII at CUTs performed as in **A)**. **E)** Heatmap analyses of CUTs performed as in **B)**.

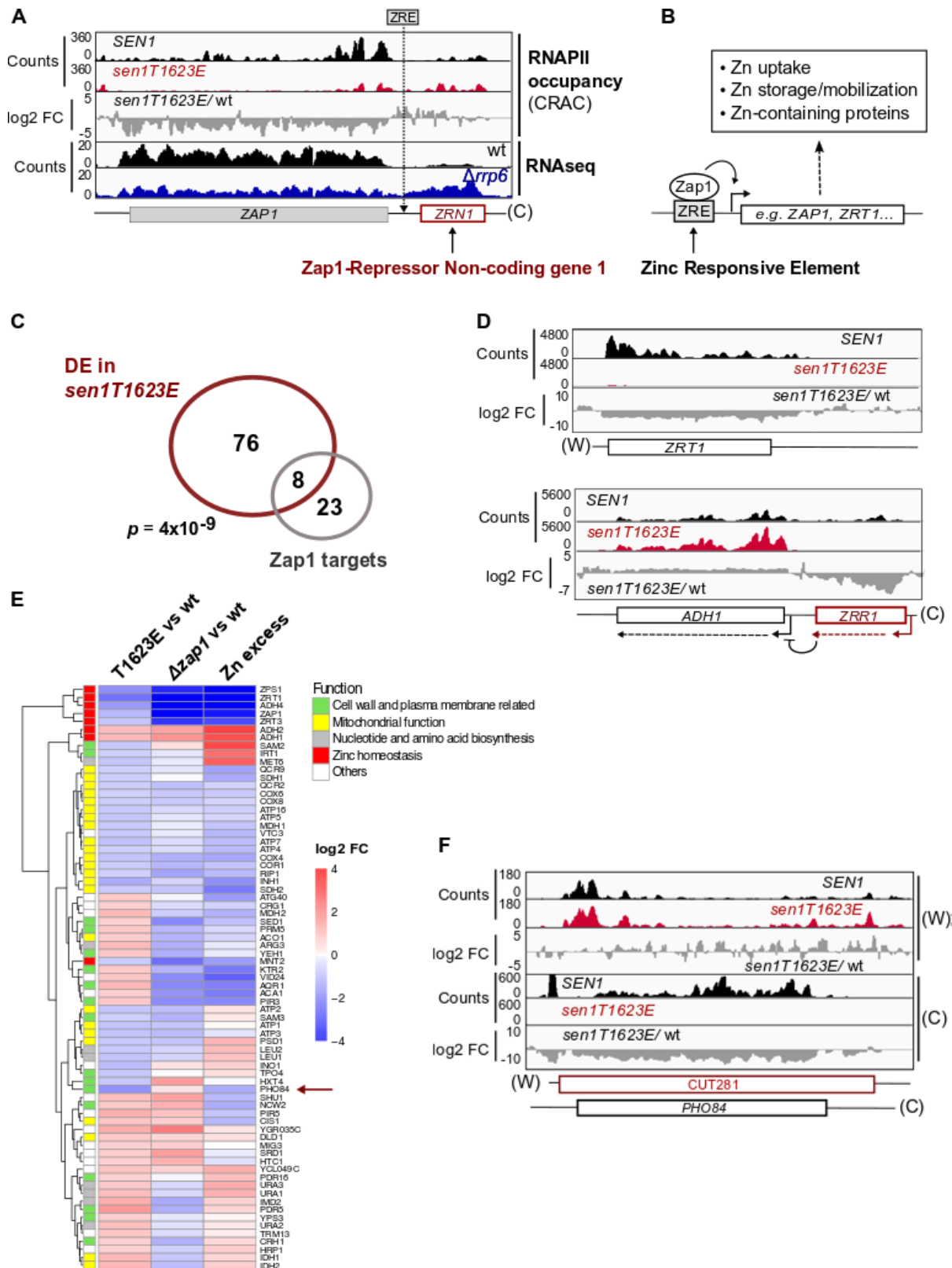
**F)** IGV screenshot of an example of CUTs exhibiting transcription termination defects in the *sen1T1623E* mutant.



**Figure 4: The T1623E mutation in Sen1 induces changes in the expression of functionally-related genes**

**A)** Analysis of differentially expressed protein-coding genes in the *sen1T1623E* mutant relative to the wt. Genes are considered as upregulated or downregulated if the log<sub>2</sub> fold change (FC) *sen1T1623E* / wt is >0.5 or <-0.5, respectively, and the p-value <0.05. The full list of genes with their corresponding expression levels and associated p-values is provided in table S2.

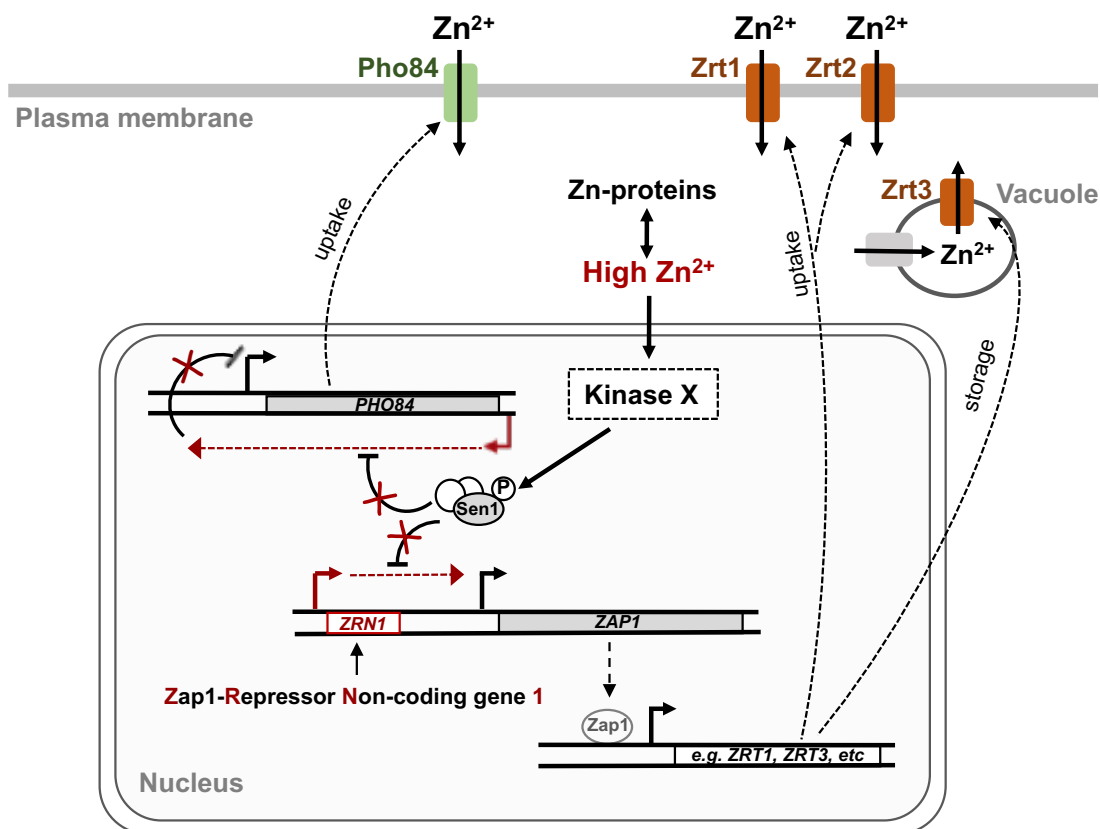
**B)** Analysis of non-coding transcription relative to genes differentially expressed (DE) in the *sen1T1623E* mutant relative to the wt. An asterisk denotes a statistically significant (p-value = 0.02) enrichment in genes transcribed immediately downstream of annotated CUTs and SUTs. Genes encoded in the mitochondrial genome were excluded from these analyses since they are transcribed by a dedicated RNAP. **C)** Representative subset of gene ontology (GO) terms found significantly overrepresented in in the *sen1T1623E* mutant relative to the wt. *p* represents the associated p-value. The full list of GO terms is provided in tables S3 and S4.



**Figure 5: Repression of *ZAP1* underlies a fraction of changes in gene expression provoked by *Sen1* T1623E mutation.**

**A)** IGV screenshot showing the *ZAP1* gene and the upstream non-coding gene *ZRN1*. The location of the previously identified ZRE at the *ZRN1-ZAP1* intergenic region is indicated. The RNAseq datasets will be published elsewhere. **B)** Schematic of the mode of action of Zap1 at regulated promoters and summary of the main processes regulated by Zap1. **C)** Venn diagram showing the overlap between the set of DE genes and Zap1 direct regulatory targets.  $p$  represents the associated p-value calculated with the Fisher's exact test. **D)** Examples of well-characterized Zap1-targets that are deregulated in the *sen1T1623E* mutant. Expression of the *ZRR1* induces repression of *ADH1* by transcriptional interference. **E)** Heatmap and hierarchical clustering analyses of the expression

profile of DE genes in the indicated datasets. The "*Δzap1* vs wt" dataset correspond to conditions of Zn starvation. "Zn excess" is an abbreviation of Zn excess vs limitation. Genes were colored according to the functional category that was assigned to them using the same color code as in figure 4E. **F)** IGV screenshot of CRAC datasets showing the *PHO84* gene and the antisense repressor non-coding gene.

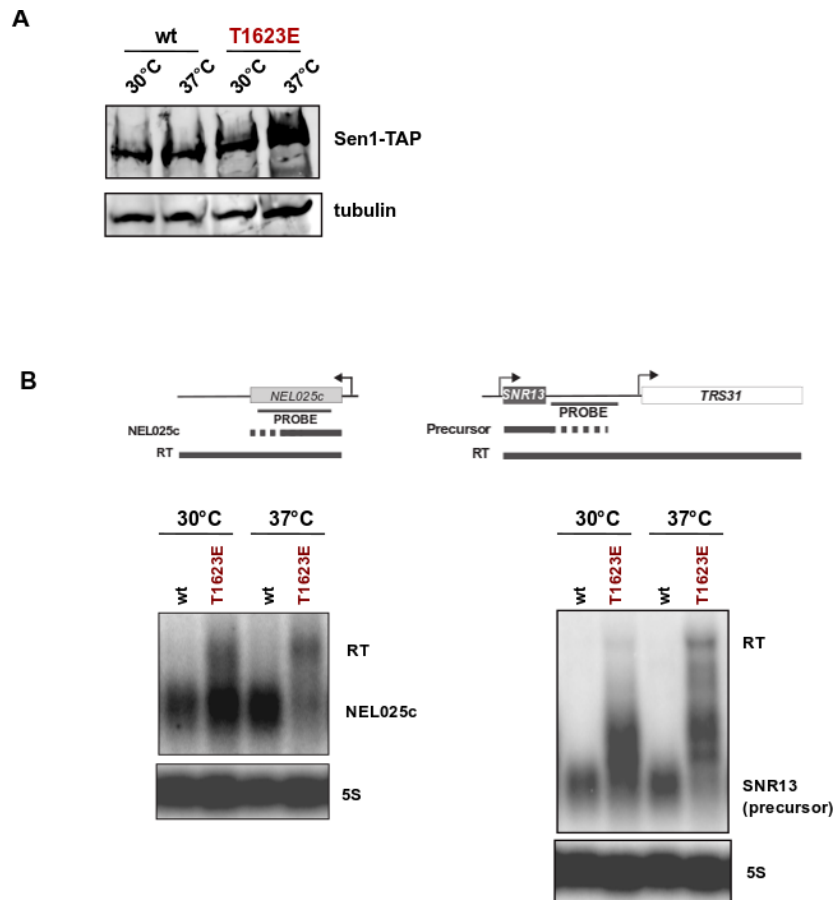


**Figure 6: Working model for the role of Sen1 phosphorylation at T1623 in the regulation of gene expression.** The NNS-complex normally terminates transcription of *ZRN1* as well as the ncRNA antisense to *PHO84*. Under Zn excess, a kinase yet to be identified would phosphorylate Sen1 to reduce the efficiency of transcription termination and promote repression of *ZAP1* and *PHO84*. Downregulation of *ZAP1* expression would lead to decreased levels of Zap1 protein and, consequently, the regulation of genes involved in Zn homeostasis, as for instance genes responsible for Zn uptake and storage and genes encoding Zn-containing proteins. Repression of *PHO84* expression, encoding a low-affinity Zn transporter, would also contribute to the regulation of Zn homeostasis.

**Table 1:** List of genes differentially expressed in the *sen1T1623E* mutant grouped by functional categories. Upregulated genes are indicated in red, whereas downregulated genes are labeled in blue.

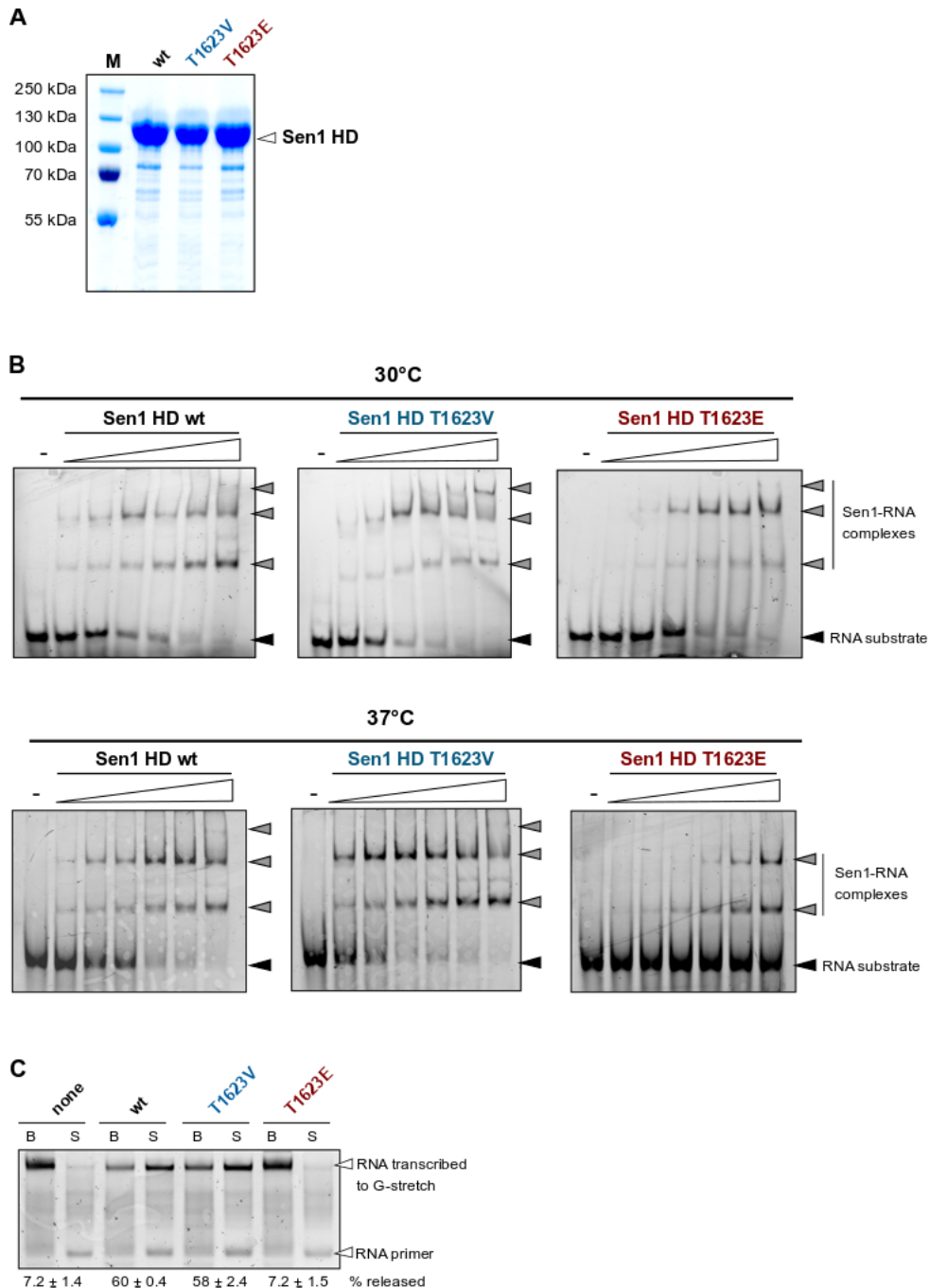
Functional category	Genes
Mitochondrial-related	<i>IDH1, IDH2, COX2, COX3, DLD1, CIS1, ACO1, ATP9, INH1, COX1, COB, ATP7, COX4, ATP1, ATP2, ATP3v, ATP5, PSD1, COR1, QCR2, MDH1, ATP4, ATP16, COX8, COX6, QCR9, RIP1, SDH2, MIN8, SDH1, ATP6</i>
Nucleotide and amino acid biosynthesis	<i>IMD2, ARG3, URA2, URA1, URA3, LEU1, LEU2, MET6</i>
Zinc homeostasis	<i>ADH1, ADH2, ZRT1, ADH4, ZPS1, ZAP1, MNT2, ZRT3</i>
Cell wall and plasma membrane-related	<i>PDR5, AQR1, SPO24, PRM5, YEH1, KTR2, NCW2, PIR3, YPS3, PDR16, CRH1, SED1, PHO84, SAM2, HXT4, SAM3, ITR1, TPO4</i>
Other functions	<i>PIR5, YGR035C, MDH2, SHU1, YCL049C, VID24, SCE1, TRM13, ATG40, SRD1, RRT15, HRP1, ACA1, HTC1, MIG3, CRG1, DPI8, INO1, VTC3</i>

## SUPPLEMENTARY FIGURES



**Figure S1: Phenotypic characterization of the *sen1T1623E* mutant at 37°C.**

**A)** SDS-PAGE analysis of C-terminally TAP-tagged Sen1 wt or T1623E expressed in yeast. Cells were grown at permissive temperature (30°C) and kept at 30°C or at high (37°C) temperature for 2h before being harvested for protein extract preparation. Proteins were detected by western blot using an antibody against the protein A moiety of the TAP-tag. Tubulin is detected as a loading control. **B)** Northern blot analysis of ncRNAs typically targeted by Sen1 for transcription termination. Cells were grown at permissive temperature (30°C) or incubated at high (37°C) temperature for 2h. Experiments were performed in a  $\Delta rrp6$  background. The 5S RNA is used as a loading control. Probes used for RNA detection are described in table S7.



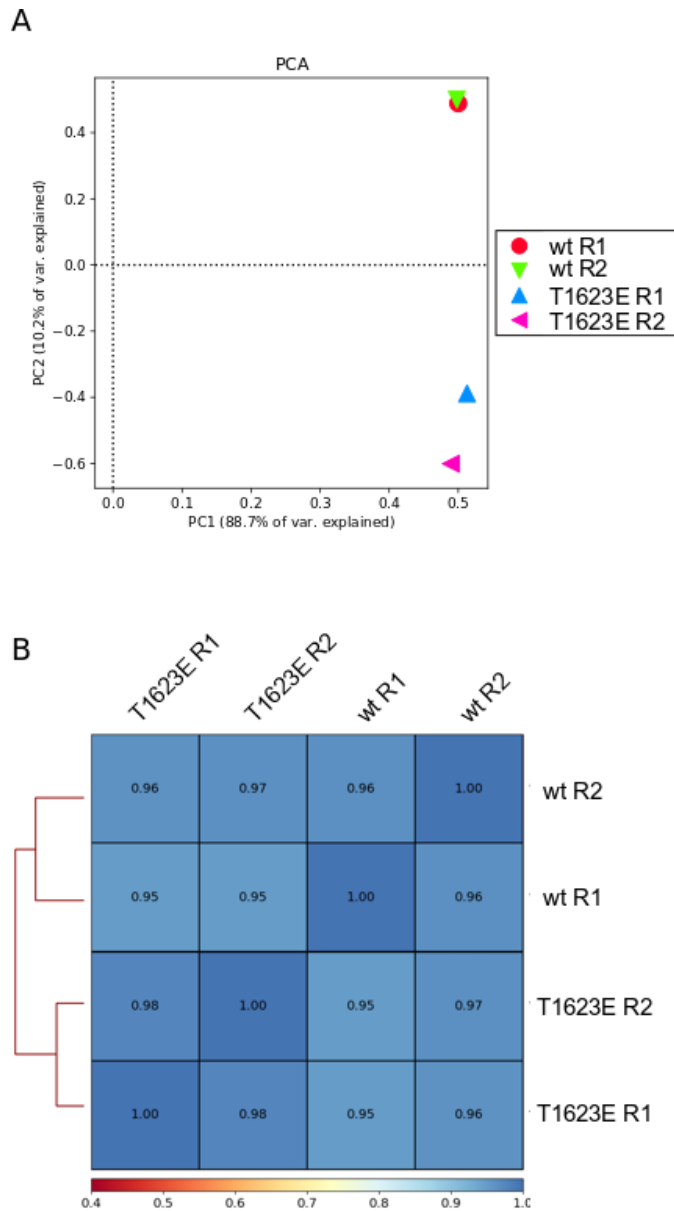
**Figure S2: Related to figure 2.**

**A)** SDS-PAGE analysis of purified Sen1 HD variants used in the experiments in figure 2.

**B)** Representative gels of EMSA summarized in figure 2C. For each experiment, the gels used for the wt, the T1623V and the T1623E mutant version of Sen1 HD were migrated and processed in parallel. A black arrowhead denotes the RNA substrate whereas the Sen1-RNA complexes are indicated by a grey arrowhead. Bands corresponding to several Sen1 HD molecules per RNA can be observed as the protein concentration increases.

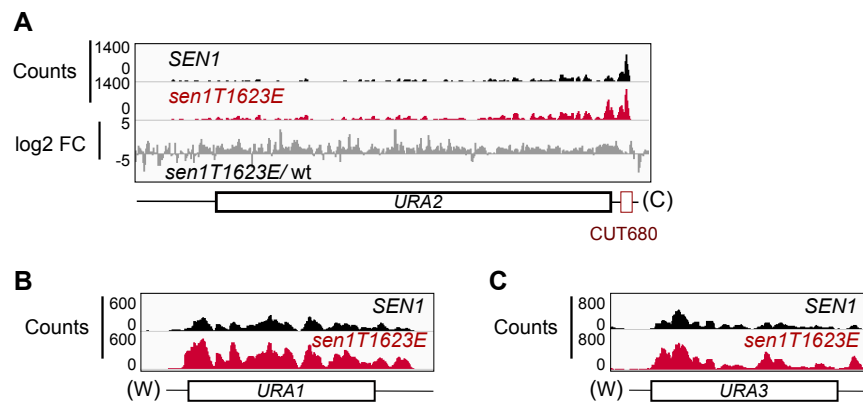
**C)** Representative gel of *In vitro* translation assays (IVTT) performed at 37°C. The values on the bottom of the gel correspond to the average and SD of three independent experiments.



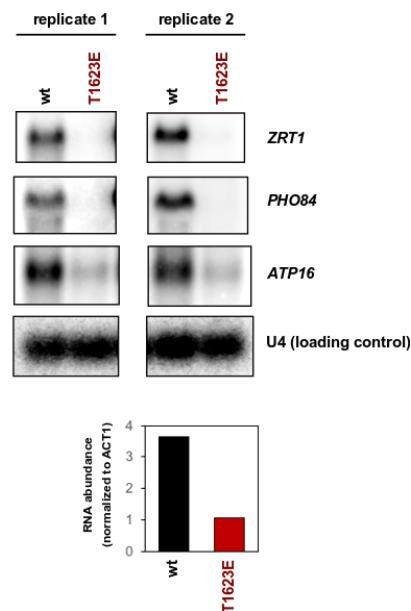


**Figure S3: Principal component analysis and correlation plots illustrating the high reproducibility of the two biological replicates of CRAC experiments in figure 3.**

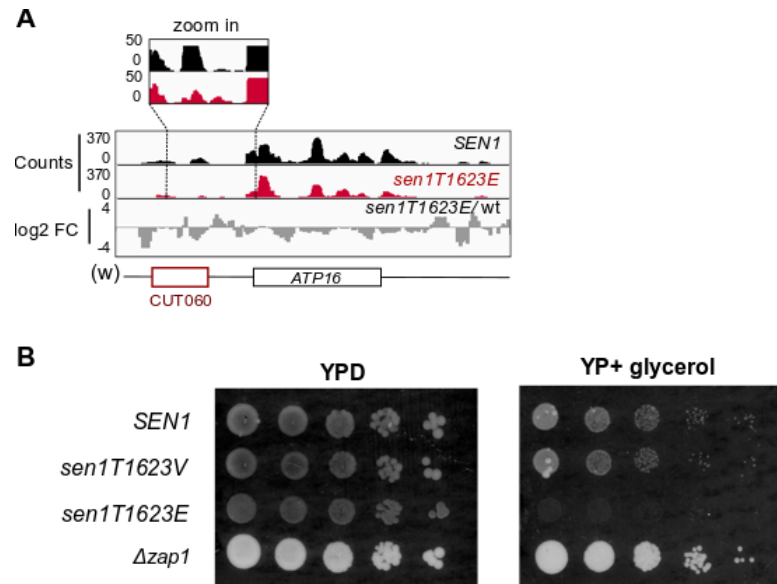
**A)** Principal component analysis (PCA) performed using the normalized reads coverage on the Watson strand (only the top 1000 bins of 500 bp). R1 and R2 denote the biological replicate 1 and 2, respectively. **B)** Heatmap representation of the correlation between the different datasets calculated as for the PCA with the coverage values of 500 bp bins in the Watson strand. The heatmap values correspond to the spearman coefficient.



**Figure S4: Deregulation of genes involved in UMP biosynthesis by the Sen1 T1623E mutation.** **A)** IGV screenshot of the *URA2* gene, likely deregulated by inefficient transcription termination at an upstream CUT. **B)** and **C)** IGV screenshots of additional genes of the UMP biosynthetic pathway that are most likely indirectly deregulated by the Sen1 T1623E mutation.



**Figure S5: Northern blot analysis of mRNAs deregulated by the Sen1 T1623E mutation.** RNA samples were prepared from aliquots of the same cultures used for CRAC experiments. The U4 RNA is used as a loading control. Probes used for RNA detection are described in table S7. Because of problems with the detection of *ZAP1* mRNA by northern blot, we show the quantification of this RNA by RT-qPCR. Values correspond to the RNA abundance relative to the *ACT1* mRNA and multiplied by 1000.



**Figure S6: Deregulation of mitochondrial-related genes by the Sen1 T1623E mutation.**

**A)** IGV screenshot of the *ATP16* gene encoding a component of the mitochondrial respiratory chain. Top: Zoom in view of the CUT060-*ATP16* intergenic region to better visualize a moderate increase in the RNAPII signal at this region in the *sen1T1623E* mutant. **B)** Growth tests in rich medium containing either glucose (YPD) or glycerol (YP + glycerol) as the carbon source.

## SUPPLEMENTARY TABLES

### Tables to be downloaded using a link:

**Table S1:** mass spectrometry analyses of Sen1-TAP.

<https://www.dropbox.com/s/rwnxtw8hwubn5fd/Table%20S1.xlsx?dl=0>

**Table S2:** results of differential expression analyses in the *sen1T1623E* relative to the wt strain.

<https://www.dropbox.com/s/v4r3kwegvz68uhrr/Table%20S2.xlsx?dl=0>

**Table S3:** List of GO terms overrepresented in the set of genes upregulated in the *sen1T1623E* mutant compared to the wt.

<https://www.dropbox.com/s/2gqvi9urk6ms3sj/Table%20S3.xlsx?dl=0>

**Table S4:** List of GO terms overrepresented in the set of genes downregulated in the *sen1T1623E* mutant compared to the wt.

<https://www.dropbox.com/s/gucd6u1s2nqzehk/Table%20S4.xlsx?dl=0>

**Table S5:** List of yeast strains used in this study.

<b>Number</b>	<b>Name</b>	<b>Genotype</b>	<b>Source</b>
DLY671	BMA	<i>as W303, Δtrp1</i>	F. Lacroute
DLY814	<i>Δrrp6</i>	<i>as W303, rrp6::KAN</i>	(Porrua et al., 2012c)
DLY2571	<i>rbp1-HTP</i>	<i>as BMA, rbp1::HTP::TRP1kl</i>	(Candelli et al, 2018)
DLY2767	<i>Δsen1/pFL38-SEN1</i>	<i>as BMA, sen1::KAN, harbouring plasmid pFL38-SEN1</i>	(Han et al, 2020)
DLY3411	<i>sen1-T1623E, rbp1-HTP</i>	<i>as BMA, sen1-T1623E, rbp1-HTP::TRP</i>	This work
DLY3340	<i>sen1-T1623E</i>	<i>as BMA, sen1-T1623E</i>	This work
DLY3345	<i>sen1-T1623E, Δrrp6</i>	<i>sen1-T1623E, rrp6::KAN</i>	This work
DLY3381	<i>sen1-T1623V</i>	<i>as BMA, sen1-T1623V</i>	This work
DLY3389	<i>sen1-T1623V, Δrrp6</i>	<i>as BMA, SEN1::sen1-T1623V, rrp6::KAN</i>	This work
DLY2131	<i>sen1-TAP</i>	<i>as BMA, SEN1::TAP::HIS</i>	This work
ZHY6	<i>Δzap1</i>	<i>as W303, ZAP1::TRP1</i>	Zhao and Eide, 1997

**Table S6:** List of plasmids used in this study.

Name	Description	Source
pDL693	Ap <sup>r</sup> ; <i>oriColE1</i> ; derivative of pFL39 bearing yeast <i>SEN1</i>	F. Lacroute
pDL772	Ap <sup>r</sup> ; <i>oriColE1</i> ; derivative of pFL38 (URA) bearing yeast <i>SEN1</i>	F. Lacroute
pDL893	Ap <sup>r</sup> ; <i>oriColE1</i> , vector for overexpression of proteins from the T7 bearing <i>sen1</i> -HD(1095-1904)CPD-His <sub>8</sub>	(Leonaite et al, 2017)
pDL921	Ap <sup>r</sup> ; <i>oriColE1</i> ; derivative of pFL39 bearing <i>sen1-T1623E</i>	This work
pDL926	Ap <sup>r</sup> ; <i>oriColE1</i> , , vector for overexpression of proteins from the T7 bearing <i>sen1-HD-T1623E-CPD- His<sub>8</sub></i>	This work
pDL940	Ap <sup>r</sup> ; <i>oriColE1</i> ,, vector for overexpression of proteins from the T7 bearing <i>sen1-HD-T1623V-CPD- His<sub>8</sub></i>	This work
pDL970	Ap <sup>r</sup> ; <i>oriColE1</i> ; derivative of pFL39 bearing <i>sen1-T1623V</i>	This work

**Table S7:** List of oligonucleotides used in this study.

Name	Sequence (5'-3')	Information/use
DL1452	CTTCCCCGTAGAAAATCTTA	Fwd primer to generate by PCR a probe to detect SNR13
DL1120	TCCGTGTCTCTTGTCTGCA	Rev primer to generate PCR a probe to detect SNR13
DL474	GCAAAGATCTGTATGAAAGG	Forward oligo to generate by PCR a probe to detect NEL025C.
DL480	ATCTGACCAGGTCAAGCTAC	Reverse oligo to generate by PCR a probe to detect NEL025C.
DL4651	TGAACCTGCCTGCTAAGTCAGG	Forward oligo to generate by PCR a probe to detect <i>ATP16</i> mRNA
DL4652	ATTGCAGCTTCTGCGGCTTC	Reverse oligo to generate by PCR a probe to detect <i>ATP16</i> mRNA
DL4688	CGGTGGTGACTACCCACTAT	Forward oligo to generate by PCR a probe to detect <i>PHO84</i> mRNA
DL4689	TGGCACCGACCTTACCAGAT	Reverse oligo to generate by PCR a probe to detect <i>PHO84</i> mRNA
DL2627	ATTCAAAAGCGAACACCGAATTGACCAT GAGGAGACGGTCTGGTTTAT	Reverse oligo used as a probe to detect U4 snRNA
DL377	ATGTTCCCAGGTATTGCCGA	Forward oligo to generate by PCR a probe to detect <i>ACT1</i> mRNA.
DL378	ACACTTGTGGTGAACGATAG	Reverse oligo to generate by PCR a probe to detect <i>ACT1</i> mRNA.
DL743	GGTTGCGGCCATATCTACCA	Reverse oligo used as a probe to detect the 5S rRNA
DL4659	GCCATGGGCCCTATGTGTTG	Forward oligo to generate by PCR a probe to detect <i>ZRT1</i> mRNA
DL4660	GCGCAGTGTAAGAACCGCTG	Reverse oligo to generate by PCR a probe to detect <i>ZRT1</i> mRNA
DL2492	UGCAUUUCGACCAGGC	5' FAM labeled RNA oligo for performing IVTT assays on immobilized templates
DL3352	CTAGAGGAAACAACTATAGGAAACGA CCAGGCCCTCAACATCTCTACCCATC TCCACACGGGGTTACCCGGCCTGCA	Non-template strand for IVTT assays
DL3353	GGCCGGGTAACCCCGTGTGGAGATG GGTGAAGATGTTGAGGGCCTGGTCGT TTCCTATAGTTTGTTCCT	Template strand for IVTT assays
DL3508	UUCAUUUCAGACCAGCACCCACUCACU ACAACUCACGACCAGGC	5' FAM labeled RNA to be used as substrate for EMSA experiments

**Addendum:**

For technical reasons we could not complete all the necessary experiments to submit this article to a scientific journal before the thesis manuscript had to be completed. The most important experiments that we expect to finish soon and that will be included in the final article are:

- Northern blot analyses in a wt and a *sen1T1623E* strains in the absence and in the presence of a plasmid expressing *ZAP1* from a heterologous promoter (already constructed) to confirm that the deregulation of zinc homeostasis genes in the *sen1T1623E* mutant is due to repression of *ZAP1* expression. Conversely, these experiments will tell whether the deregulation of *PHO84* and *ATP16* genes in *sen1T1623E* is independent on the repression of *ZAP1*, as we propose.
- Northern blot analyses in a wt and a *sen1T1623V* strain in the presence of different concentrations of Zn (i.e. from Zn excess to limitation) to test whether in the strain expressing the T1623V version of Sen1, which cannot be phosphorylated, *ZAP1* expression responds differently to Zn concentration. If *ZAP1* is partially derepressed under Zn excess in the *sen1T1623V* mutant, this would support our model proposing that the phosphorylation of Sen1 T1623 plays a role in the regulation of *ZAP1* expression when the concentration of Zn is high.

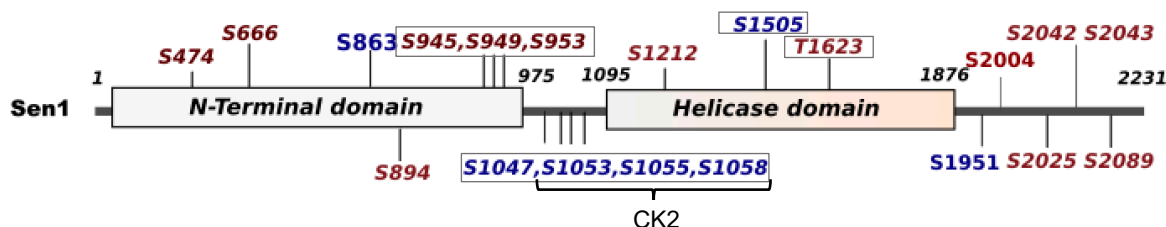


### **3.2 Analyse globale du rôle des phosphorylations de Sen1 dans la réponse au stress**

### 3.2.1 Identification des sites de phosphorylation de la protéine Sen1 :

Comme mentionné ci-dessus, l'hélicase Sen1 est la sous-unité catalytique du complexe NNS. Notre hypothèse est que la plupart des régulations pourraient s'opérer sur Sen1 soit au niveau de son recrutement et/ou au niveau de son activité catalytique, éventuellement par des modifications post-traductionnelles. Pour tester cette hypothèse, nous avons réalisé une expérience d'immunoprécipitation de Sen1 suivie par une analyse par la spectrométrie de masse.

Cette expérience a permis d'identifier un certain nombre de partenaires de Sen1 parmi lesquels figure l'ARN polymérase III. La caractérisation du rôle de cette interaction est en cours au sein de notre laboratoire. Un article dans lequel je figurerai comme 4<sup>ème</sup> auteur sur ce projet est en cours de préparation. Quant aux modifications post-traductionnelles, nous avons identifié 13 sites de phosphorylations au sein de la protéine Sen1 (figure 22). Parmi les phosphosites de Sen1, nous avons identifié une séquence consensus phosphorylée par la caséine kinase 2 (CK2), une kinase qui joue différents rôles dans la transcription (Meggio and Pinna, 2003).



**Figure 22 :** Inventaire des sites de phosphorylations dans la protéine Sen1.

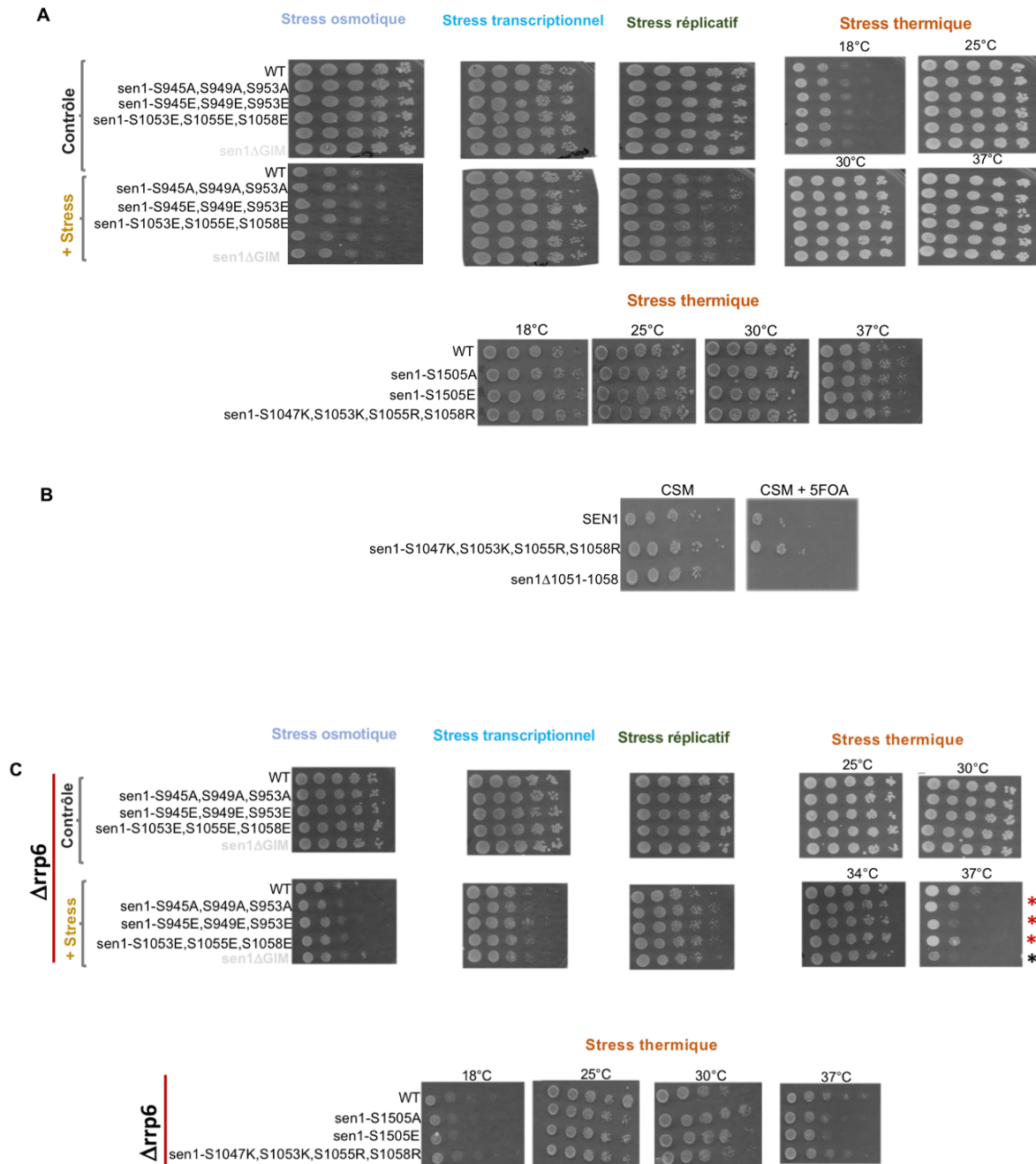
Les phosphosites identifiés durant ma thèse sont colorés en rouge tandis que les phosphosites précédemment décrits dans la littérature sont colorés en bleu (MacGilvray et al., 2020; Danielle L Swaney et al., 2013). Enfin, les sites encadrés sont ceux que nous avons tentés de caractériser au cours de ma thèse.

Afin d'aborder la régulation de Sen1 par la phosphorylation, nous nous sommes tout d'abord focalisés à la fois sur des phosphosites qui sont localisés dans les domaines fonctionnels essentiels de Sen1 et qui apparaissent comme une cible présumée de kinase. En effet, comme mentionné ci-dessus, le domaine N-terminal de Sen1 est essentiel à la fois pour la viabilité cellulaire et la terminaison de la transcription non-codante, nous avons tenté d'analyser le rôle de certaines phosphorylations identifiées dans cette région, plus précisément les phosphosites S945, S949, et S953. Quant aux phosphosites situés entre le domaine N-terminal et le domaine hélicase, nous avons imaginé que sa phosphorylation par CK2 pourrait participer à la modulation des interactions intramoléculaires au sein de Sen1. Dans notre précédente étude, nous avons montré dans un système hétérologue que le domaine N-terminal de Sen1 interagit avec la région C-terminal (Han et al., 2020). Il est possible d'imaginer que ces phosphorylations, localisées dans la région charnière entre ces deux domaines, pourraient être impliquées dans cette interaction. Enfin, les phosphorylations localisées dans le domaine hélicase qui porte l'activité catalytique de Sen1 ont aussi attiré notre attention. La phosphorylation de la S1505 nous a paru potentiellement intéressante car elle est localisée dans un sous-domaine nommé le (« prong ») essentiel pour la viabilité cellulaire et la terminaison de la transcription non-codante (Leonaitè et al., 2017). Quant au phosphosite T1623, sa caractérisation est largement présentée à travers l'article de thèse.

### 3.2.2 Analyse du rôle des phosphorylations de Sen1 dans la croissance cellulaire :

Afin de caractériser le rôle des phosphosites localisés dans le domaine N-terminal, nous avons introduit concomitamment une triple mutation au niveau des positions S945, S949, S953. Ces résidus ont été remplacés par le glutamate ou l'alanine afin de mimer ou d'empêcher les phosphorylations. Dans le cas des phosphorylations des sites putatives de CK2, nous avons introduit une triple mutation qui mime ou qui présente un effet opposé à la phosphorylation (mutation vers la lysine et l'arginine) ou encore nous avons totalement délété totalement cette séquence. Enfin, la position 1505 du domaine hélicase a subi une mutation qui mime ou qui empêche la phosphorylation. Pour l'ensemble de ces sites, nous avons procédé d'abord à une analyse de l'effet de ces mutations sur la croissance cellulaire dans les conditions normales et dans différentes conditions de stress (stress thermique, choc osmotique, stress transcriptionnel et le stress réplicatif). Cette analyse a révélé que les mutants ponctuels ne sont pas affectés dans la croissance (figure 23A) mais la délétion du site putative de CK2 conduit à la létalité suggérant l'importance de cette région dans le repliement ou le fonctionnement de la protéine (figure 23B).

En revanche, lorsque ces mutations ont été combinées à la délétion de *RRP6* (une des sous-unités majeures de l'exosome nucléaire qui dégrade les ARNs cibles du complexe NNS) certains de ces mutants présentent un défaut de croissance à 37°C plus prononcé excepté le mutant de la sérine en position 1505. De manière intéressante, le mutant phosphomimétique S945, S949, S953 présente un défaut de croissance plus exacerbé comparé au mutant de l'alanine suggérant que la protéine Sen1 pourrait être partiellement inhibée par la phosphorylation de ces résidus (figure 23C). De plus, certains mutants affectés dans la terminaison de la transcription dépendante du complexe NNS révèlent souvent une interaction génétique avec le mutant  $\Delta rrp6$ . Cela nous a amené à imaginer que ces phosphorylations pourraient réguler l'activité de Sen1 et par conséquent, moduler l'efficacité de la terminaison de la transcription et, éventuellement, participer à la régulation de l'expression des gènes.



**Figure 23** : Analyse de la croissance de phospho-mutants de Sen1 dans plusieurs conditions.

Le stress osmotique, transcriptionnel et réplcatif ont été induits respectivement avec 1M de NaCl, 1,5% de formamide et 100 mM d'hydroxyurée. Quant au stress thermique, les cellules ont été incubées à différentes températures, 30°C étant la température normale de croissance. Les phospho-mutants dans un fond génétique sauvage (A et B) ou dans un fond  $\Delta rrp6$  (C) sont testés dans les différentes conditions indiquées. B : La souche  $\Delta sen1$  contenant un plasmide *URA3* (pFL38) exprimant *SEN1* a été transformée avec un plasmide *TRP1* (pFL39) exprimant soit la version sauvage de *SEN1* ou la version mutante. Après une croissance nocturne en milieu non sélectif, les cellules ont été étalées sur un milieu minimum (CSM) contenant de l'acide 5-fluorootique (5-FOA) pour sélectionner celles qui ont perdu le plasmide *URA3*. L'absence de croissance dans le milieu contenant 5-FOA implique que la version de *sen1* exprimée dans le plasmide *TRP1* ne restaure pas la viabilité. Les boîtes sont photographiées au bout de 2 à 4 jours. L'astérisque rouge correspond aux seules conditions où des défauts de croissance sont observés. L'astérisque noir correspond à un mutant qui sera discuté dans la section 3.3 des résultats de ce manuscrit.

### 3.2.3 Analyse du rôle des phosphorylations de Sen1 dans la terminaison de la transcription non-codante :

Ensuite, j'ai entrepris d'analyser l'effet de ces mutations sur la terminaison de la transcription non-codante par northern blot. Les mutations sont combinées à  $\Delta rrp6$  afin de rendre les ARN non-codants détectables. L'analyse de l'efficacité de la terminaison de deux cibles classiques (NEL025C et SNR13) de la voie NNS dans les conditions normales de croissance a révélé qu'aucun des mutants ne présentent un défaut de terminaison sur ces deux gènes (données non montrées).

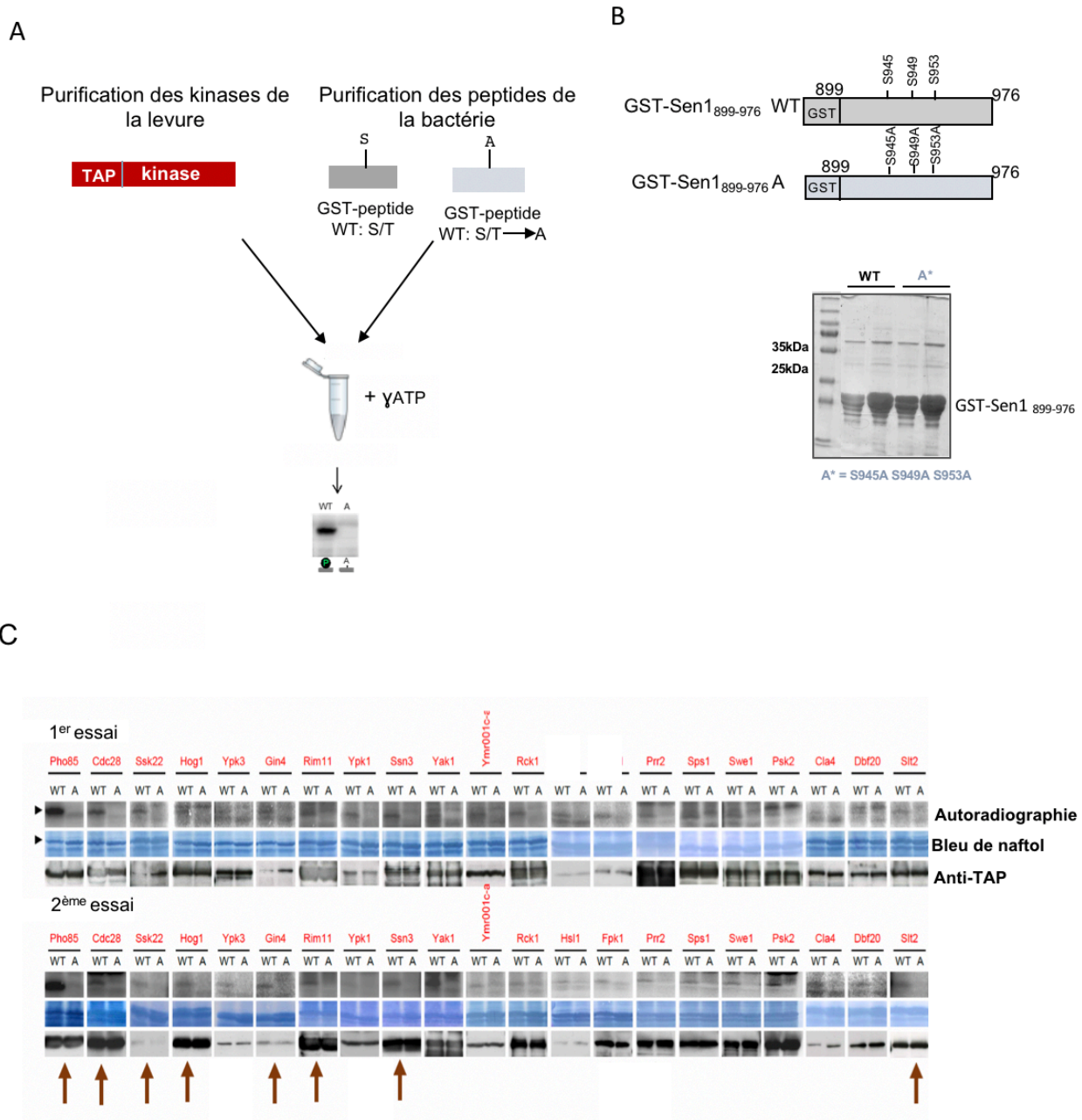
Le défaut de croissance plus marqué du mutant phosphomimétique S945E, S949E, S953E dans le fond  $\Delta rrp6$  nous a conduit à se focaliser sur ce mutant et analyser si ce phénotype est lié à une diminution de l'efficacité de la terminaison non-codante à 37°C. Ainsi, nous avons réalisé une analyse des ARNs extraits de cellules incubées 2H à 37°C. Ces analyses n'ont révélé aucun défaut de terminaison relatifs aux deux cibles classiques de la voie NNS testées (données non montrées) indiquant que ces phosphorylations ne jouent pas un rôle général dans la terminaison de la transcription non-codante dépendant du complexe NNS. Néanmoins, Il est possible d'imaginer que ces mutations pourraient affecter seulement une partie des gènes non-codants ciblés par la voie NNS et/ou affecter l'expression de gènes qui sont régulés par le complexe NNS à travers une terminaison précoce.

Pour tester cette possibilité, nous avons réalisé des analyses transcriptomiques (RNAseq) dans le mutant phosphomimétique combiné aux mutations  $\Delta rrp6$ ,  $\Delta Upf1$  car les ARN non-codants cibles de la voie NNS qui échappent à la dégradation nucléaire sont stabilisés dans le mutant  $\Delta Upf1$  et deviennent ainsi détectables. Ces analyses ont révélé que ces phosphorylations ne modulent pas la terminaison par le complexe NNS dans les conditions de croissance normale. En revanche, il ne permet pas d'exclure que ces phosphorylations puissent avoir un rôle dans des conditions physiologiques particulières. Par exemple, ces modifications pourraient affecter l'efficacité de la terminaison de la transcription non-codante et contribuer à réguler l'expression des genes dans certaines conditions (e.g. conditions de stress). En ce sens, un exemple très original de régulation de l'expression par la transcription d'un ARN non-codant a été décrit par le laboratoire de Francesc Posas (IRB, Barcelone, Espagne). Il s'agit du gène *CDC28* codant pour une kinase impliquée dans le cycle cellulaire. L'induction de ce gène par un ARN non-codant en antisens permet à la cellule de redémarrer rapidement le cycle cellulaire après le stress. Dans la continuité de la même étude, les auteurs ont montré que 68 CUTs (cibles classiques de la voie NNS) sont induits ou stabilisés en réponse au stress osmotique (Nadal-Ribelles et al., 2014). La stabilisation de ces ARN non-codants pourrait être liée à un défaut de terminaison de la transcription en raison d'un dysfonctionnement de la voie NNS. De ce fait, les ARN non-codants sont pris en charge par la voie CPF-CF et échappent à la dégradation nucléaire. En vue d'identifier la condition physiologique dans laquelle des phosphorylations dans Sen1 pourraient participer à la modulation de la terminaison de la transcription, nous avons cherché à identifier les kinases responsables de ces modifications.

### 3.2.4 Identification des kinases responsables de la phosphorylation des sérines 945, 949 et 953

Pour tenter d'identifier les kinases responsables de la phosphorylation des sérines 945, 949 et 953, nous avons entrepris une collaboration avec l'équipe de F. Posas pour réaliser un criblage *in vitro* avec une banque de kinases purifiées de la levure à travers un test de phosphorylation *in vitro* (figure 24A). Pour résumer, nous avons surexprimé chez *E. coli* un fragment de Sen1 contenant les résidus d'intérêt ainsi qu'une version mutée non-phosphorylable. Les deux polypeptides sont fusionnés à une étiquette GST (glutathion-S-transférase). Nous avons purifié ces fragments par chromatographie d'affinité (figure 24B) et nos collaborateurs les ont incubés avec les différentes kinases purifiées de la levure en présence de l'ATP radio-marqué sur le phosphate gamma. Une détection radioactive a permis ainsi de révéler la phosphorylation des substrats. Les kinases qui se révèlent capables de

phosphoryler ledit fragment de Sen1 sauvage mais pas la version mutante dans deux réactions indépendantes sont retenues comme des kinases qui pourraient potentiellement phosphoryler Sen1 *in vivo* sur une ou plusieurs de ces sérines. Ce criblage nous a permis d'identifier les kinases Pho85, Cdc28, Ssk22, Hog1, Gin4, Rim11, Ssn3 et Slt2 comme capables de phosphoryler Sen1 *in vitro* (figure 24C).



**Figure 24 :** Identification des kinases capables de phosphoryler les sérines 945, 949 et/ou 953 de Sen1 *in vitro*. A : Le système expérimental simplifié est illustré selon ce schéma. B : Le schéma des fragments de Sen1 purifiées est indiqué (en haut) et les protéines sont révélées après coloration au bleu de coomassie. C : Les résultats présentés ici correspondent à celles qui ont pu phosphoryler Sen1 lors d'un premier 1<sup>er</sup> essai et qui ont été re-testées (2<sup>ème</sup> essai). Les flèches indiquent les kinases pour lesquels le signal semble être plus clair et reproductible.

Afin d'identifier la condition physiologique dans laquelle la protéine Sen1 pourrait être régulée par phosphorylation, après une revue de littérature sur chacune de ces kinases capables de phosphoryler Sen1 *in vitro*, nous avons pu constater que la plupart de ces kinases sont impliquées dans la réponse au stress. Parmi elles, la kinase Hog1, le régulateur central de la réponse au stress osmotique a attiré

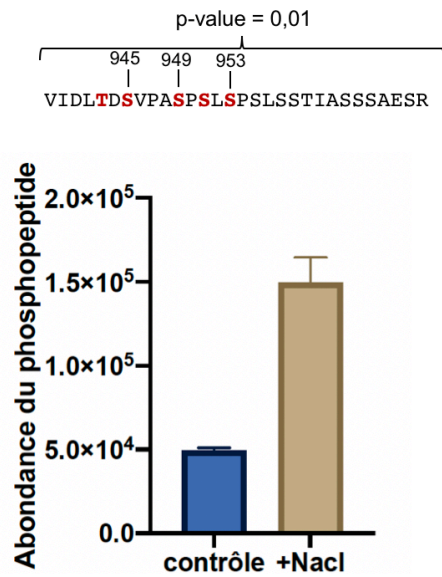
notre attention car il a été révélé que Hog1 est capable d'induire l'expression d'environ 200 ARN non-codants en réponse au stress osmotique (Nadal-Ribelles et al., 2014). Dans la continuité de la même étude, les auteurs ont pu révéler qu'une partie importante de ces ARN non-codants chevauchent avec les CUTs et les SUTs qui sont dans leur grande majorité métabolisés par la voie NNS. En outre, la transcription de certains de ces genes non-codants serait impliquée dans la régulation de l'expression des genes voisins. Cela nous a conduit à imaginer que la phosphorylation de Sen1 par Hog1 pourrait participer à la régulation de l'expression des genes en réponse au stress osmotique via la modulation de la terminaison par Sen1.

### **3.2.5 Analyse de la régulation de Sen1 en réponse de stress osmotique :**

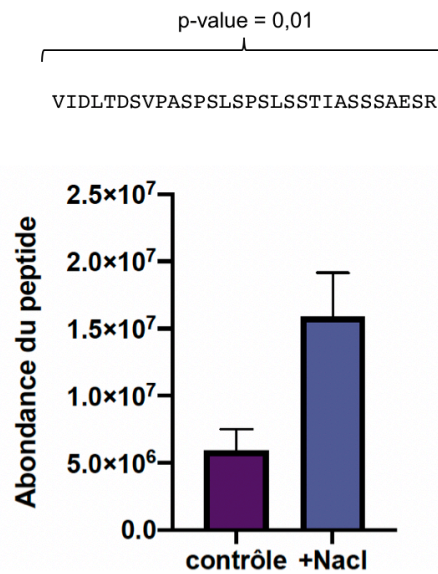
#### **3.2.5.1 Analyse de l'état de phosphorylation de Sen1 en condition de stress :**

Afin de tester si Sen1 est en effet phosphorylée dans les sérines 945, 949, 953 en réponse au stress osmotique, nous avons réalisé une expérience d'immunoprécipitation suivie par une analyse par spectrométrie de masse (MS) avec quantification sans marquage afin d'analyser l'état de phosphorylation de Sen1 en condition de stress osmotique. Pour induire ce stress, nous avons incubé les cultures en présence de NaCl 0,4 M pendant 10 minutes ou en présence d'un volume équivalent d'eau comme contrôle. Cette analyse a permis de révéler en condition de stress une augmentation de la quantité de peptide de Sen1 contenant les sites putatifs de phosphorylation par Hog1, aussi bien dans la forme phosphorylée que non-phosphorylée (figure 25A-B). En conséquence, ce résultat ne permet pas de conclure d'une augmentation de la phosphorylation de Sen1 en réponse au stress osmotique et ouvre des questions sur la stratégie expérimentale utilisée. En effet, la MS semble être une technique limitée pour comparer le niveau de phosphorylations des protéines dans deux conditions physiologiques pour la raison suivante : la forme phosphorylée de chaque peptide par rapport à la forme non phosphorylée est généralement faible à la détection car les liaisons phosphates sont labiles et peuvent se casser au cours de l'analyse MS/MS et les formes phosphorylées s'ionisent généralement moins bien que les formes non phosphorylées. Néanmoins, Il serait intéressant de procéder à un enrichissement des phosphopeptides sur une colonne de dioxyde de titane avant de procéder à l'analyse et d'utiliser un mutant  $\Delta hog1$  afin de mieux évaluer l'état de phosphorylation de Sen1 en condition de stress osmotique ainsi que le rôle de Hog1 dans la phosphorylation de Sen1 dans cette condition. Une autre alternative s'avère être intéressante : le système phos-tag qui est une technologie innovante pour analyser l'état de phosphorylation des protéines. Elle repose sur une molécule conçue pour piéger les phosphoprotéines de manière spécifiques. Après migrations des protéines sur gel, les protéines phosphorylées sont détectées à l'aide d'un anticorps et présentent une différence de taille sur gel par rapport à la forme non-phosphorylée. Cependant, cette technique présente des limitations quand il y a des nombreuses phosphorylations dans la protéine d'intérêt et seulement une petite fraction de ces phosphorylations change dans les conditions d'étude, ce qui pourrait être le cas de Sen1. La grande taille de la protéine Sen1 est aussi une contrainte pour ce type d'approches.

A



B



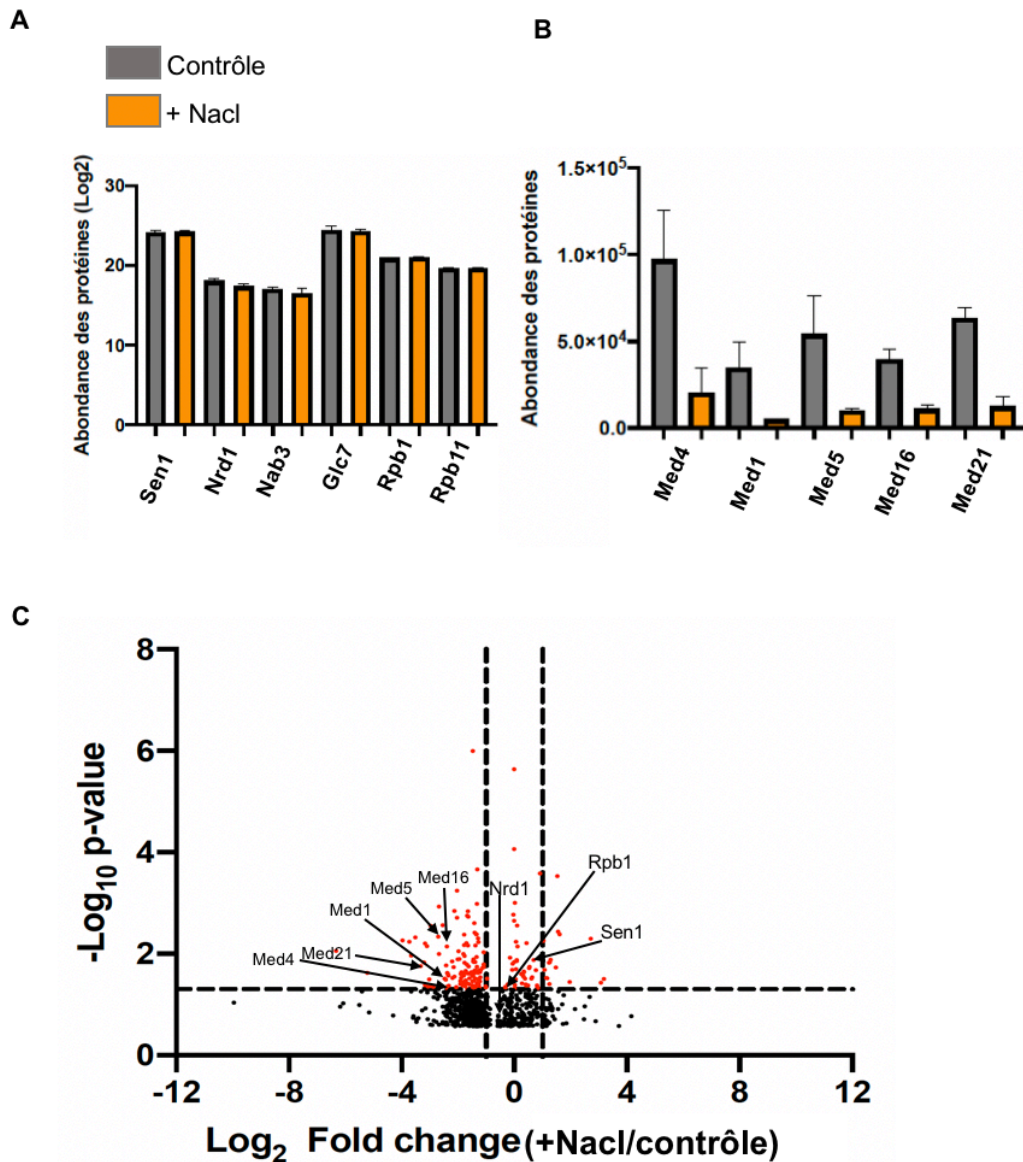
**Figure 25 :** Analyse du niveau de phosphorylation de Sen1 en condition normale vs stress osmotique

A : L'histogramme représente la quantité de phosphopeptides. B : L'histogramme indique la quantité de la forme non phosphorylée du peptide. En haut de chaque histogramme, les peptides quantifiés contenant les sites putatifs de Hog1. La p-value correspond à la probabilité que la quantité de peptides soit égale entre les deux conditions analysées.

### 3.2.5.2 Analyse du réseau d'interaction de Sen1 en condition de stress :

La même expérience d'immunoprécipitation suivie par une analyse à la spectrométrie de masse en condition de stress osmotique nous a permis d'analyser la façon dont le réseau d'interaction de Sen1 pourrait être modifié dans cette condition. Pour cela, nous avons comparé la condition de croissance normale au stress osmotique. Les analyses réalisées montrent que les niveaux d'expression de Sen1 ainsi que son interaction avec ses partenaires traditionnels ne semblent pas être affectés (figure 26A). En revanche, de manière très intéressante nous avons observé une diminution de l'interaction de Sen1 avec le complexe « Mediator » (figure 26B). Ce complexe est une plateforme de recrutement essentiel pour la transcription de tous les gènes et sert d'interface physique entre les facteurs de transcription et le complexe d'initiation de la transcription. Il a été proposé que l'état de phosphorylation du complexe « Mediator » change en condition de stress osmotique (Miller et al., 2012). Il est donc tentant de spéculer que la modulation de la phosphorylation du « Mediator » en condition de stress pourrait réguler son interaction avec la protéine Sen1.





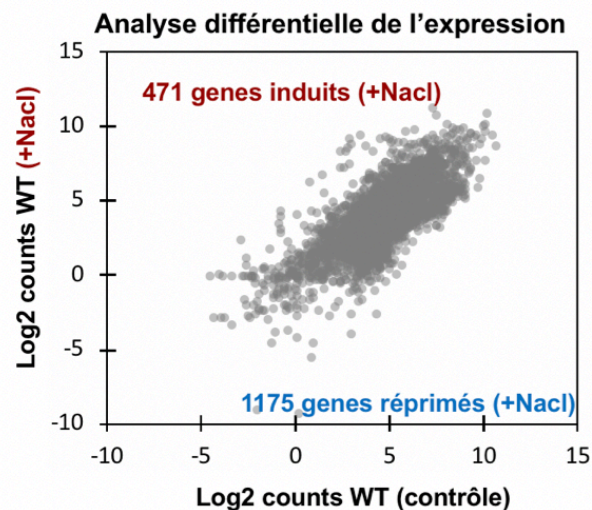
**Figure 26** : Identification des partenaires de Sen1 par co-immunoprécipitation suivi par une analyse par spectrométrie de masse sans marquage. Après identification, les protéines ont été ainsi quantifiées. A : L'histogramme représente l'interaction de Sen1 avec ses partenaires traditionnels. Les protéines Rpb1 et Rpb11 sont des sous unités de l'ARNpol II. B : L'histogramme indique l'interaction de Sen1 avec le « Mediator » C : Volcano plot illustrant les interactions affectées en condition de stress osmotique. La ligne discontinue correspond à la p-value = 0,05

### 3.2.6 Analyse à haute résolution de la transcription de l'ARNpol II en condition de stress osmotique:

L'exploration du rôle possible de ces phosphorylations de Sen1 dans la réponse au stress osmotique nous a conduit à réaliser une analyse transcriptomique dans une souche sauvage et une souche *sen1 S945A, S949A, S953A* (en bref, *S945A*) cultivées en milieu minimal avec ou sans traitement au NaCl 0,4 M pour induire le stress osmotique. À cette fin, nous avons utilisé une technique appelée CRAC, mentionnée dans la section précédente, qui permet de cartographier l'ARN naissant lors de son association avec l'ARNpol II au cours de la transcription. Ces expériences étaient également intéressantes indépendamment de la fonction des phosphosites de Sen1 car les analyses du transcriptome en condition de stress osmotique précédant ne reposaient que sur la détection d'ARN stables, et ne pouvaient donc révéler qu'une petite fraction du transcriptome non-codante (Nadal-

Ribelles et al., 2014). En permettant la détection de tous les ARNs transcrits, quelle que soit leur stabilité, la technique CRAC offrirait une vision inédite du paysage transcriptionnel en condition de stress osmotique et pourrait potentiellement mettre en évidence de nouveaux événements régulateurs participant à la réponse au stress.

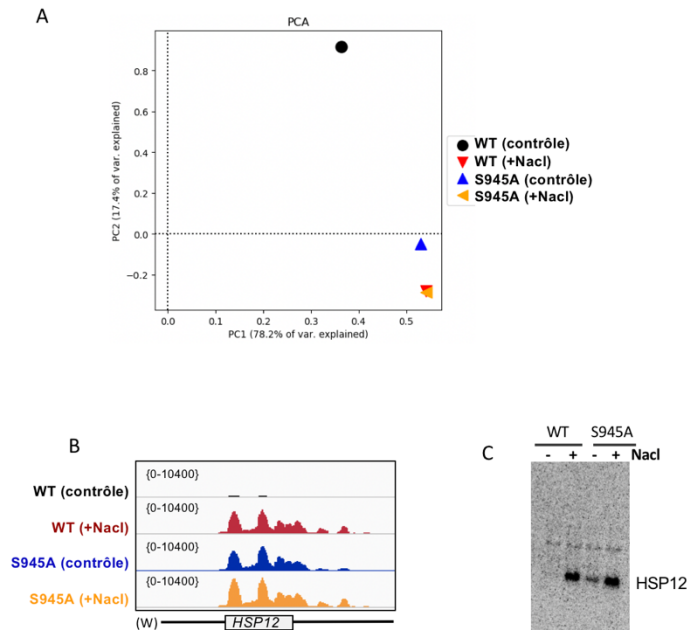
Des analyses préliminaires d'expression différentielle effectuées avec une seule réplique biologique pour chaque condition ont révélé des changements dans l'expression d'environ 25 % des gènes codant pour des protéines (figure 27). Ces changements ont consisté à l'activation de près de 500 gènes, dont de nombreux gènes impliqués dans la réponse au stress, ainsi qu'en la répression de plus de 1000 gènes, dont la plupart sont impliqués dans la prolifération cellulaire. Ces changements sont en accord avec les rapports précédents (revue : Clotet and Posas, 2007; Yaakov et al., 2009) et valident nos conditions expérimentales.



**Figure 27 :** Analyse préliminaire de l'expression différentielle effectuée sur les gènes codant les protéines.

Le signal ARNpol II total recouvrant les gènes codant les protéines ( $n=6601$ ) a été calculé et comparé dans les deux conditions. Les gènes ont été considérés comme induits en condition de stress osmotique (+NaCl) lorsque le signal est 2 fois supérieur et considérés comme réprimés lorsque ce signal est 2 fois inférieur par rapport au groupe contrôle. Les analyses de « gene ontology » ont révélé un enrichissement des gènes de réponse au stress parmi les gènes induits (par exemple, 6,8 % des gènes impliqués dans le processus de réponse au stress oxydatif,  $p\text{-value}=2,9 \times 10^{-9}$ ) et un enrichissement des gènes impliqués dans la traduction et la prolifération parmi les gènes réprimés (par exemple, 23,8 % des gènes impliqués dans la biogenèse des ribosomes,  $p\text{-value}=7,1 \times 10^{-104}$ ), comme attendu.

Une analyse globale de l'expression des gènes de la souche sauvage comparé au mutant indique un profil d'expression des gènes très similaire en condition de stress (figure 28A). Ce résultat nous fournit deux explications possibles : i) la phosphorylation de Sen1 par Hog1 observé *in vitro* ne se produit pas *in vivo* ii) Hog1 pourrait phosphoryler Sen1 sur des résidus supplémentaires et la mutation de l'ensemble de sites de phosphorylation serait nécessaire pour observer un phénotype. De manière surprenante, le gène *HSP12*, l'un des marqueurs clés de l'induction au stress osmotique et dépendant de Hog1 (de Nadal et al., 2004) est induit dans les deux souches en condition de stress, comme attendu, mais aussi dans le mutant S945A en absence de stress (figure 28B-C). De plus, le mutant S945A présente dans la condition contrôle un profil d'expression qui ressemble partiellement au profil en condition de stress comme si la simple mutation suffisait pour induire un stress (Figure 28A). Ce résultat peut suggérer un artefact expérimental et une répétition de l'expérience sera essentielle pour déterminer la reproductibilité de ces observations.

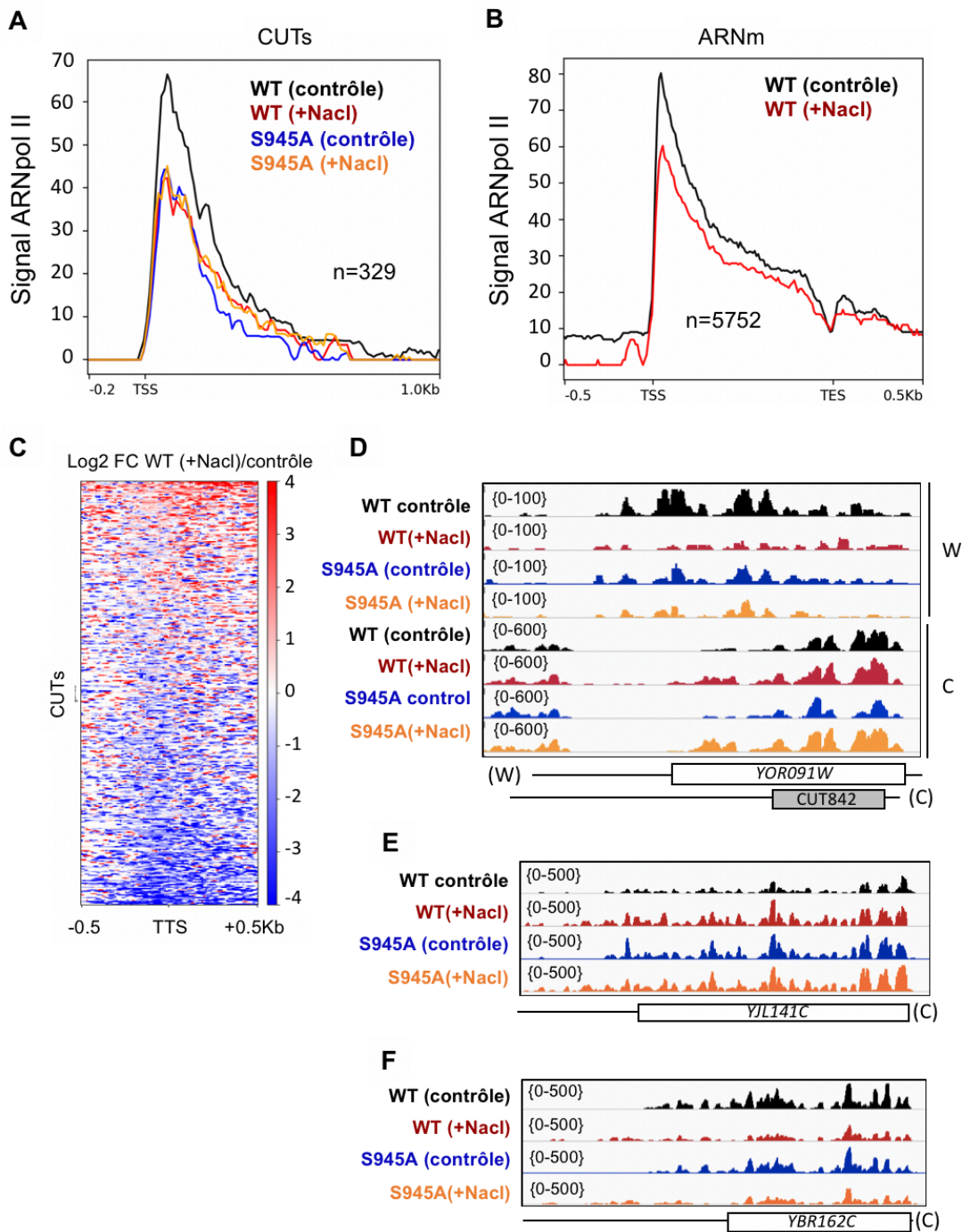


**Figure 28** : Analyse du profil d'expression du mutant S945A par rapport au WT.

A : Analyse en composantes principales effectuée sur les niveaux d'expression des gènes codant pour les protéines. B : Capture d'écran de l'IGV montrant le signal ARNpol II au niveau du gène *HSP12*, qui est typiquement activé sous divers stress, dont le stress osmotique. C : analyses par Northern blot d'ARNs préparés à partir des mêmes cultures que celles utilisées pour les expériences de CRAC.

Les analyses « metagène » du signal ARNpol II au niveau de CUTs nous ont fourni des observations intéressantes. Alors que le niveau d'ARNpol II moyen a diminué sous l'effet du stress à la fois au niveau des CUT et au niveau des ARNm (figure 29A-B), ce qui est en accord avec une diminution globale de la transcription, cette diminution a progressivement disparu vers le 3' des CUTs, ce qui pourrait indiquer des défauts de terminaison dans certains cas. Une analyse plus détaillée par « heatmap » a confirmé cette hypothèse en révélant une augmentation du signal ARNpol II en aval du site de terminaison annoté pour une fraction de CUTs (figure 29C). Il est intéressant de noter que dans certains cas, ces défauts de terminaison au niveau des CUTs étaient accompagnés de la répression d'un gène codant une protéine en antisens, ce qui suggère que la modulation négative de la terminaison dépendante de Sen1 pourrait participer à la régulation de certains gènes en réponse au stress osmotique (figure 29D). Une identification et une évaluation plus précises des CUT affectées dans la terminaison de la transcription ainsi que leur implication dans la régulation des gènes nécessiteront au moins une répétition supplémentaire de ces expériences.

Il est intéressant de noter que nous avons observé un comportement similaire lorsque nous avons analysé les ARNm. En effet, certains gènes codant de protéines présentant des défauts clairs de terminaison de la transcription en condition de stress osmotique (figure 29E-F). La pertinence biologique de ce phénomène ainsi que la question de savoir si ce mécanisme participe à la régulation de l'expression des gènes restent des questions intéressantes à explorer pour les études futures.



**Figure 29** : Analyse à haute résolution de la transcription par l'ARN Pol II en condition de stress osmotique par le CRAC.

A et B : Analyses « metagène » du signal de l'ARNpol II aux CUTs et aux ARNm. Les valeurs correspondent à la couverture par la médiane. C : Heatmap représentant le ratio du log2 du changement du signal ARNpol II (FC pour "fold change") sur les CUTs dans la souche sauvage en condition de stress osmotique par rapport au contrôle. Le TSS désigne le site d'initiation de la transcription (« termination start site ») alors que le TTS désigne le site de terminaison de la transcription (« transcription termination site »). D, E et F : Les captures d'écran à partir du logiciel IGV illustrent les CUTs et les gènes codant des protéines présentant des défauts de terminaison de la transcription en condition de stress osmotique.

### 3.2.7 Résumé et perspectives :

Dans ce travail, nous avons d'abord identifié plusieurs sites de phosphorylation dans la protéine Sen1 par une expérience d'immunoprécipitation suivie par une analyse à la spectrométrie de masse. Sur les phosphosites les plus pertinents, nous avons procédé à l'analyse du rôle des phosphorylations sur la croissance cellulaire en condition de stress et sur l'efficacité de la terminaison de la transcription non-codante. Nous nous sommes concentrés sur les phosphorylations des positions 945, 949 et 953 car le positionnement de ces phosphosites dans le domaine N-terminal de Sen1, une interface d'interactions protéiques, nous a conduit à formuler l'hypothèse selon laquelle ces phosphorylations pourraient moduler l'interaction de Sen1 avec d'autres facteurs et par conséquent contribuer à l'efficacité de la terminaison de la transcription non-codante. Une mutation phosphomimétique de ces résidus dans le mutant *Δrrp6* a révélé une interaction génétique avec ce mutant, ce qui pouvait être en accord avec un effet inhibiteur possible de ces phosphorylations. Cependant, nos analyses transcriptomiques effectués dans le mutant phosphomimétique dans les conditions de croissance normale ont dévoilé un paysage transcriptionnel assez similaire à celui de la souche sauvage. Ces données indiquent que ces phosphorylations ne semblent pas jouer un rôle important dans les conditions normales de croissance sans exclure le rôle de ces modifications dans des conditions physiologiques spécifiques. A travers un criblage *in vitro* avec une banque de kinases purifiées de la levure, nous avons découvert que ces résidus sont phosphorylés par un certain nombre de kinases parmi lesquelles figure Hog1, le régulateur central de la réponse au stress osmotique. Cela nous a conduit à privilégier le stress osmotique comme la condition physiologique dans laquelle Sen1 serait régulée par phosphorylation, dans le cas particulier par Hog1.

Nous avons réalisé des analyses transcriptomiques en condition de stress osmotique par la méthode CRAC en comparant la souche sauvage à la mutante (S945A, S949A, S953A) pour déterminer l'effet de la phosphorylation dans ces résidus dans la réponse au stress. Ces analyses ont révélé que le profil d'expression des gènes du mutant est similaire à celui de la souche sauvage suggérant ainsi que les phosphorylations identifiées ne sont pas importantes pour la régulation de l'expression génique dans ces conditions. Il est possible d'imaginer que des phosphorylations supplémentaires seraient nécessaires à la régulation de Sen1 en réponse au stress osmotique. Néanmoins, cette expérience nous a permis de révéler des défauts de terminaison à la fois de certains CUTs et de certains gènes codant des protéines et qui est indépendant de ces phosphorylations. Il serait intéressant d'explorer l'importance biologique de ces défauts de terminaison, ainsi que les mécanismes sous-jacents.

Enfin, nos analyses protéomiques ont révélé, de manière intéressante, une diminution en condition de stress de l'interaction de Sen1 avec le complexe « Mediator », un acteur clé de l'initiation de la transcription. En intégrant l'ensemble de ces données bien que préliminaires, il est tentant de spéculer que le défaut de terminaison observé en condition de stress pourrait être provoqué par un défaut de recrutement de Sen1 au niveau certains ARN non-codants par le complexe médiateur. Pour compléter ces données, une fois ces résultats reproduits, il serait nécessaire de réaliser les analyses bio-informatiques afin d'identifier les caractéristiques communes des ARN non-codants affectés dans la terminaison. Ensuite, on pourrait analyser si le défaut de terminaison observé en condition de stress est associé à un défaut de recrutement de Sen1 en regardant la liaison de Sen1 aux gènes affectés en conditions de stress, par exemple par la technique de CRAC. Ensuite, il serait intéressant d'explorer les mécanismes de recrutement de Sen1 par le « Mediator » et comprendre comment le recrutement de Sen1 pourrait être modulé en condition de stress. Une fois tous ces aspects décortiqués, on serait en mesure de bloquer la modulation du recrutement de Sen1 et d'évaluer ainsi la contribution réelle de ce mécanisme à la régulation de l'expression des gènes en réponse au stress osmotique.

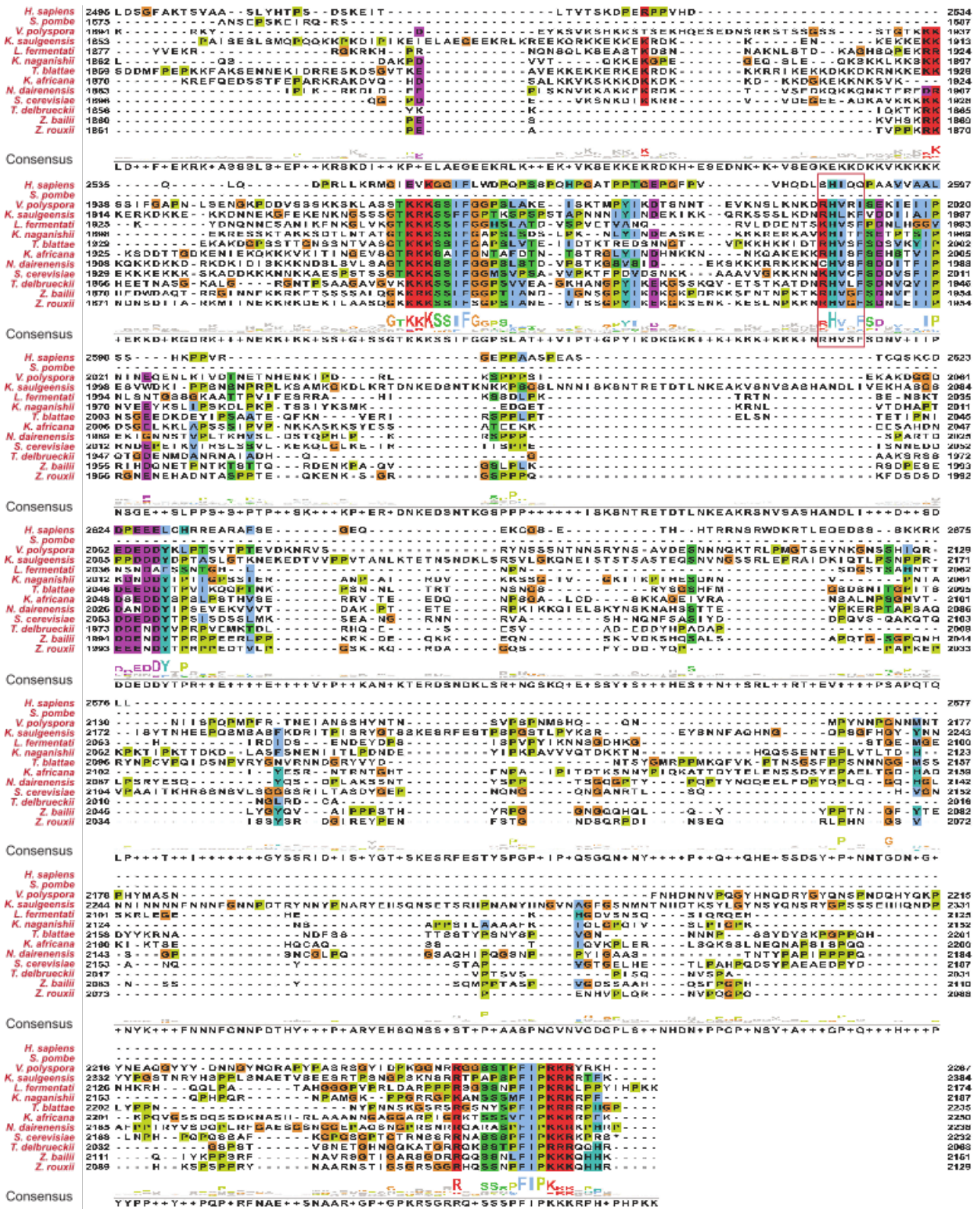
### 3.3 Analyse fonctionnelle de l'interaction de Sen1 avec la phosphatase Glc7

### 3.3.1 La phosphatase Glc7 interagit avec la région C-terminale de Sen1 :

Les phosphorylations et les déphosphorylations sont des modifications chimiques portées sur des molécules afin de moduler leur fonctionnement. Ces modifications sont au centre de tous les processus cellulaires. Les enzymes portant ces fonctions doivent être précises dans leurs actions. C'est le cas de la phosphatase Glc7, une enzyme hautement conservée. Elle reconnaît ses partenaires à travers un motif de séquence d'acides aminés dégénérés, communément simplifié sous le nom de RVxF. Ce motif est typiquement flanqué en N-terminal de résidus basiques et en C-terminal de résidus acides, et la conservation de la séquence et les analyses de mutation ont établi l'importance de cette séquence (Wakula et al., 2003). Des études précédentes ont permis d'identifier la présence de ce motif dans la région C-terminale de Sen1 et ont montré son implication dans l'interaction de Sen1 avec Glc7 (Nedeja et al., 2008).

Une comparaison de la région C-terminale de Sen1 chez différentes espèces eucaryotes a permis de révéler le degré de conservation de ce motif dans des levures proches (figure 30) que nous avons appelé GIM pour « Glc7 interaction motif ». La conservation de ce motif pourrait indiquer que l'interaction de Sen1 avec Glc7 joue un rôle important dans la fonction de Sen1.

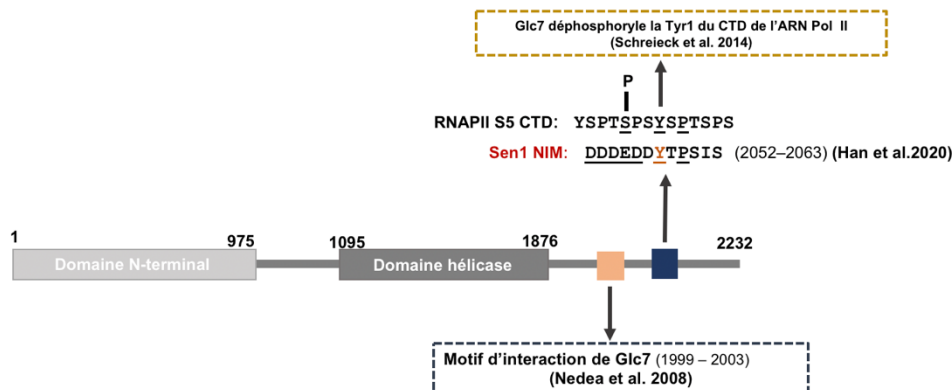
La protéine Glc7 joue un rôle dans la terminaison de la transcription des gènes codants les protéines ainsi que certains gènes non-codants (Collin et al., 2019). Elle participe à cette étape de la transcription en déphosphorylant la tyrosine 1 du CTD de l'ARNpol II et cette action semble être nécessaire pour le recrutement des facteurs impliqués dans la terminaison (Schrieck et al., 2014). Puisque Glc7 interagit avec le facteur clé de la terminaison non-codante Sen1, il est donc possible d'imaginer qu'elle pourrait aussi jouer un rôle dans la terminaison par Sen1. Comme mentionné dans l'introduction, nous avons identifié et montré à travers une étude dans laquelle j'ai participé que la région C-terminale de Sen1 contient à proximité du site d'interaction de Glc7 un motif conservé connu sous l'appellation NIM (Nrd1-Interaction Motif) qui présente une similarité avec le CTD de l'ARNpol II (figure 31) et qui est impliqué dans l'interaction de Sen1 avec Nrd1 (Han et al., 2020). Il a été montré par des approches *in vitro* que la phosphorylation de la tyrosine 1 du CTD de l'ARNpol II empêche le recrutement de Nrd1 (Mayer et al. 2010). Puisque Glc7 déphosphoryle la tyrosine 1 du CTD de l'ARNpol II, un résidu qui est aussi conservé dans le NIM de Sen1, notre hypothèse est que Glc7 pourrait déphosphoryler ce résidu afin de moduler l'interaction de Sen1 avec Nrd1.



**Figure 30 :** Le GIM est conservé dans les orthologues de Sen1  
 La séquence de la protéine Sen1 de *Saccharomyces cerevisiae* a été soumise à un blastp excluant le genre *Saccharomyces* de la recherche. Les dix séquences protéiques les plus conservées ainsi que les orthologues de Sen1 de *Homo sapiens* (SETX) et *Schizosaccharomyces pombe* ont été alignées sur Sen1 de *S. cerevisiae* en utilisant le programme omega clustal. La visualisation de l'alignement et le calcul des séquences consensus enregistrés sont effectués avec Jalview (Waterhouse et al., 2009). Les acides aminés sont colorés selon le schéma de couleur du clustal. Les identifications des protéines utilisées dans l'alignement sont les suivants : Q7Z333 (*H. sapiens*), Q92355 (*S. pombe*), XP\_001644478.1 (*Vanderwaltozyma polyspora*), SMN21961.1 (*Kazachstania saulgeensis*), SCV99407.1 (*Lachancea fermentati*), CCK69072.1 (*Kazachstania naganishii*),



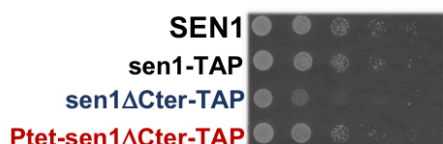
XP\_004181671.1 (*Tetrapisipora blattae*), XP\_003955130.1 (*Kazachstania Africana*), XP\_003672383.1 (*Naumovozya dairenensis*), Q00416 (*S. cerevisiae*), XP\_003680903.1 (*Torulaspota delbrueckii*), CDF91445.1 (*Zygosaccharomyces bailii*) et GAV52597.1 (*Zygosaccharomyces rouxii*). Le GIM est indiqué par un encadré rouge.



**Figure 31** : Illustration de la séquence de Sen1 avec une focalisation sur les interactions médiées par la région C-terminale de Sen1. La séquence du CTD de l'ARNpol II phosphorylé sur la sérine 5 (S5-CTD), et du NIM de Sen1 est indiquée en haut. Les résidus importants pour l'interaction avec la Nrd1 CID (CTD-Interaction Domain) sont soulignés. La tyrosine (Y2058) du NIM de Sen1, l'équivalent de la tyrosine 1 du CTD, est indiquée en orange et pourrait correspondre à la cible putative de Glc7.

### 3.3.2 Analyse de l'effet de la délétion du GIM sur la croissance cellulaire :

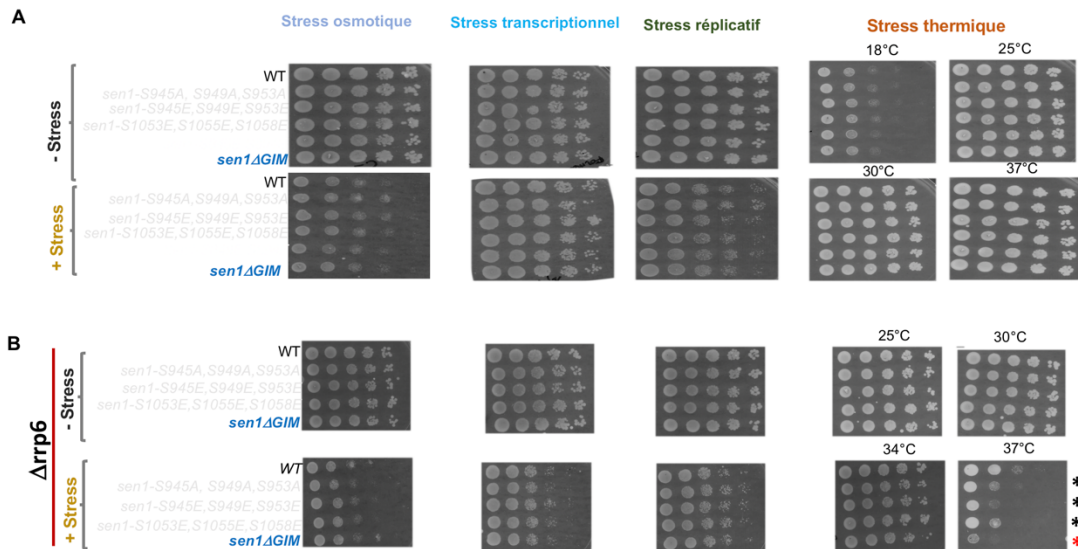
Afin de caractériser le rôle de l'interaction de Sen1 avec Glc7, nous avons tout d'abord délété totalement la région C-terminale et analysé l'effet de cette mutation sur la croissance cellulaire. De manière intéressante, nous avons montré que cette mutation conduit à des défauts de croissance mais ce défaut est supprimé par la surexpression de *sen1ΔCter* (figure 32). Par ailleurs, nous avons montré dans une étude précédente que la région C-terminale de Sen1 participe modestement à l'efficacité de la terminaison de la transcription non-codante (Han et al., 2020a). De plus, sa délétion ne conduit pas à une déstabilisation de la protéine Sen1. Ces données suggèrent que la région C-terminale de Sen1 pourrait participer à son recrutement.



**Figure 32** : la surexpression de *sen1ΔCter* supprime le défaut de croissance associé à la délétion de la région C-terminale. Le test de croissance est réalisé sur YPD et les boîtes sont incubées à 30°C puis photographié trois jours après incubation. La variante Ptet-*sen1ΔCter* correspond à une surexpression à partir du promoteur régulé par la tétracycline (Ptet) au locus endogène.

Cependant, en plus du GIM, la région C-terminale contient d'autres séquences qui pourraient contribuer au recrutement de Sen1 et au phénotype observé, notamment le NIM. Afin d'explorer le rôle précis du GIM, nous avons délété ce motif et analysé l'effet de cette délétion sur la croissance

cellulaire dans différentes conditions, y compris des conditions de stress. Ces analyses ont montré que le mutant *sen1ΔGIM* n'est pas affecté dans la croissance dans aucune des conditions testées (figure 33). En revanche, cette mutation combinée à *Δrrp6* présente un défaut de croissance plus prononcé à 37°C. Cette interaction synthétique pourrait suggérer un défaut de terminaison de la transcription non-codante.



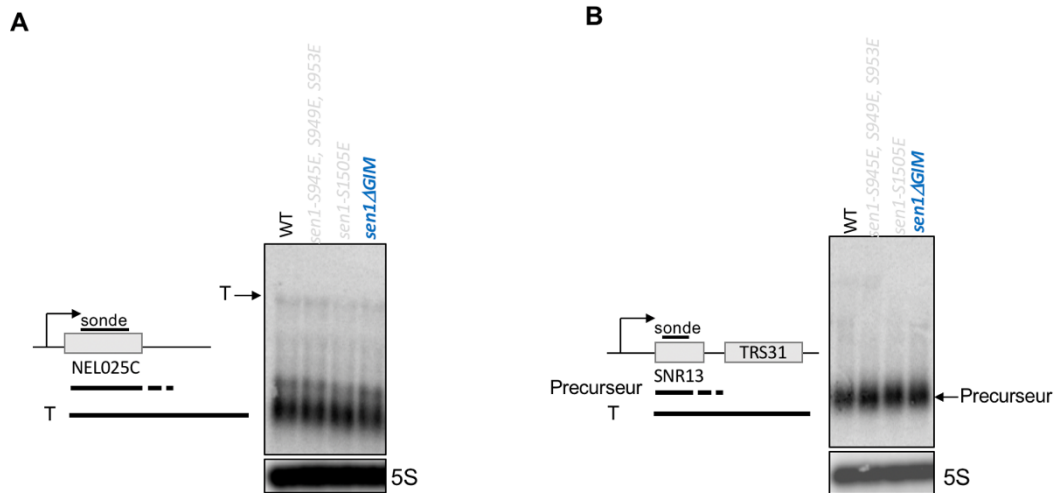
**Figure 33 :** la délétion de GIM combinée à celle de *RRP6* est associée à un modeste défaut de croissance à 37°C.

Le stress osmotique, transcriptionnel et réplcatif ont été induits respectivement avec 1M de NaCl, 1,5% de formamide et 100mM de l'hydroxyurée. Quant au stress thermique, les boites ont été incubées à différentes températures. J'ai réalisé un contrôle (-stress) pour chaque condition en ajoutant le volume d'eau correspondants au volume de drogue pour mimer la répartition des nutriments. Le panel A correspond au test de croissance réalisé dans le mutant *sen1ΔGIM* tandis que le panel B correspond à cette mutation combinée à *Δrrp6*. Les boites sont photographiées après de 2 à 4 jours d'incubation. L'astérisque en rouge correspond à la seule condition dans laquelle j'ai observée un effet sur la croissance.

### 3.3.3 Analyse du rôle de l'interaction Sen1–Glc7 dans la terminaison de la transcription non-codante:

Afin de tester le rôle de cette interaction dans la terminaison de la transcription non-codante, nous avons analysé l'impact de la délétion du GIM dans le mutant *Δrrp6, Δupf1* sur la terminaison de deux cibles classiques (NEL025C et SNR13) de la voie NNS. La délétion de *UPF1* offre un avantage supplémentaire car les ARN non-codants qui échappent à la dégradation nucléaire sont stabilisés dans ce mutant et deviennent donc détectables.

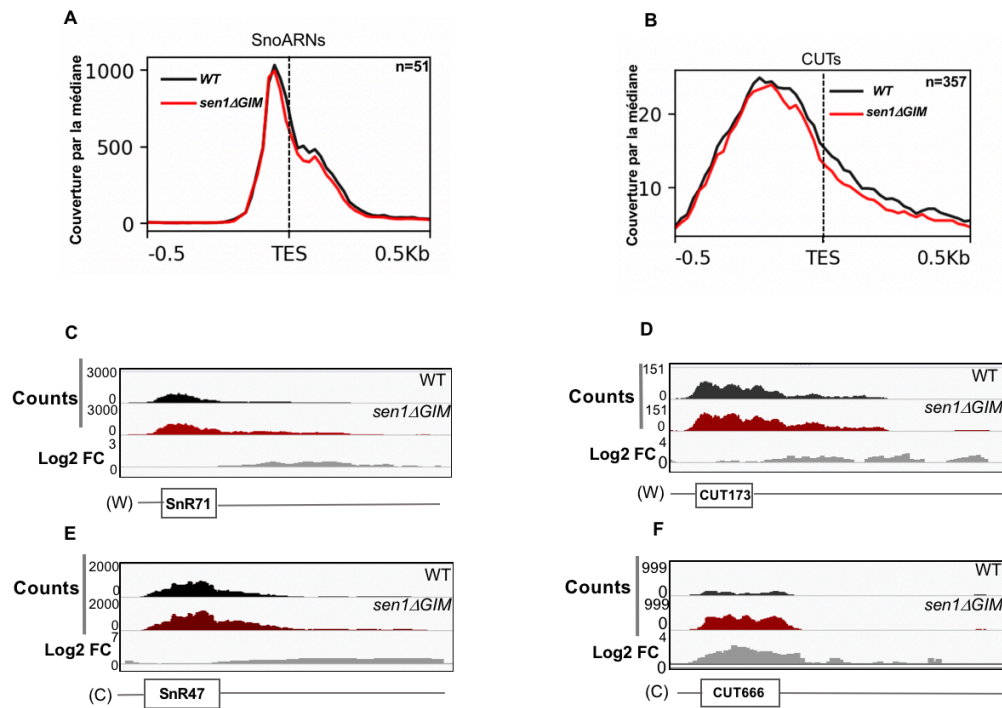
Ces analyses n'ont pas révélé de défaut de terminaison pour NEL025C et SNR13 (figure 34A-B) suggérant que l'interaction de Sen1 avec Glc7 ne joue pas un rôle majeur dans la terminaison de la transcription non-codante. Les mêmes analyses réalisées à 37°C n'ont pas révélé de défaut de terminaison (données non montrées) suggérant ainsi que le défaut de croissance observé dans ces conditions ne semble pas être lié à un défaut de terminaison.



**Figure 34 :** La délétion du GIM ne provoque pas un défaut de terminaison de la transcription de cibles classiques de la voie NNS.

Analyse de l'efficacité de la terminaison de la transcription par northern blot de NEL025C (A) et SNR13 (B) dans le mutant *sen1ΔGIM*. L'espèce longue des ARN correspondant à une translecture (T) n'est pas détectée. L'ARN ribosomal 5S est utilisé comme contrôle de charge.

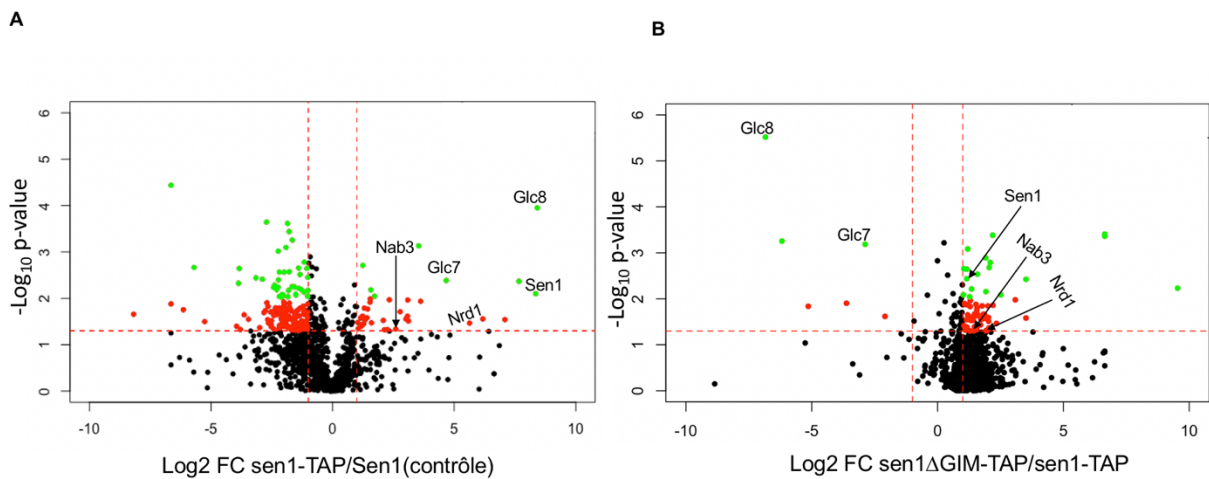
L'interaction de Sen1 avec Glc7 pourrait jouer un rôle dans la terminaison de la transcription des certains gènes non-codants de manière spécifique. Pour explorer cette possibilité, nous avons effectué des analyses transcriptomiques (RNAseq) permettant d'analyser globalement le rôle de cette interaction dans la terminaison. Ces analyses ont révélé que les CUTs et les SnoARNs ne sont pas globalement affectés dans la terminaison (figure 35A-B). En revanche, certains SnoARNs et CUTs sont modestement affectés dans la terminaison (figure 35C-D-E-F). L'ensemble de ces données suggèrent que cette interaction ne joue pas un rôle global dans la terminaison non-codante mais pourrait participer à la régulation de la terminaison de certains ARN non-codants de manière spécifique.



**Figure 35 :** L'interaction Sen1 – Glc7 contribue modestement à l'efficacité de la terminaison de la transcription  
 A et B : Analyse metagénomique des snoARNs et CUTs. Le point de référence TES est le site de terminaison de la transcription (TES pour « transcription end site »). C, D, E et F : Capture d'écran des données RNAseq à partir d'IGV de quelques exemples de CUTs et de snoARNs affectés dans la terminaison de la transcription.

### 3.3.4 Analyse du réseau d'interaction de Sen1 et de son état de phosphorylation dans le mutant *sen1ΔGIM* :

Afin d'analyser le réseau d'interaction de Sen1 et de son état de phosphorylation, nous avons réalisé une expérience de co-immunoprécipitation suivie par une analyse par la spectrométrie de masse (MS) en comparant la souche sauvage au mutant *sen1ΔGIM*. Dans la souche sauvage, cette expérience nous a permis, comme attendu, d'identifier les partenaires traditionnels de Sen1 notamment Nrd1, Nab3, Glc7 et Glc8 (figure 36A). La protéine Glc8 est une chaperonne qui interagit avec Glc7 afin de moduler son activité dans certaines conditions physiologiques (Nigavekar et al., 2002). Dans le mutant *sen1ΔGIM*, comme attendu, nous avons observé une perte de l'interaction de Sen1 avec Glc7 confirmant que le motif GIM est nécessaire pour cette interaction. De manière inattendue, nous avons observé une augmentation de l'interaction de Sen1 avec Nrd1 et Nab3 (figure 35B) suggérant le rôle de Glc7 dans la modulation de l'interaction de Sen1 avec les deux autres sous-unités du complexe NNS. Cependant, l'effet observé est l'opposé de l'effet attendu d'après notre hypothèse initiale.

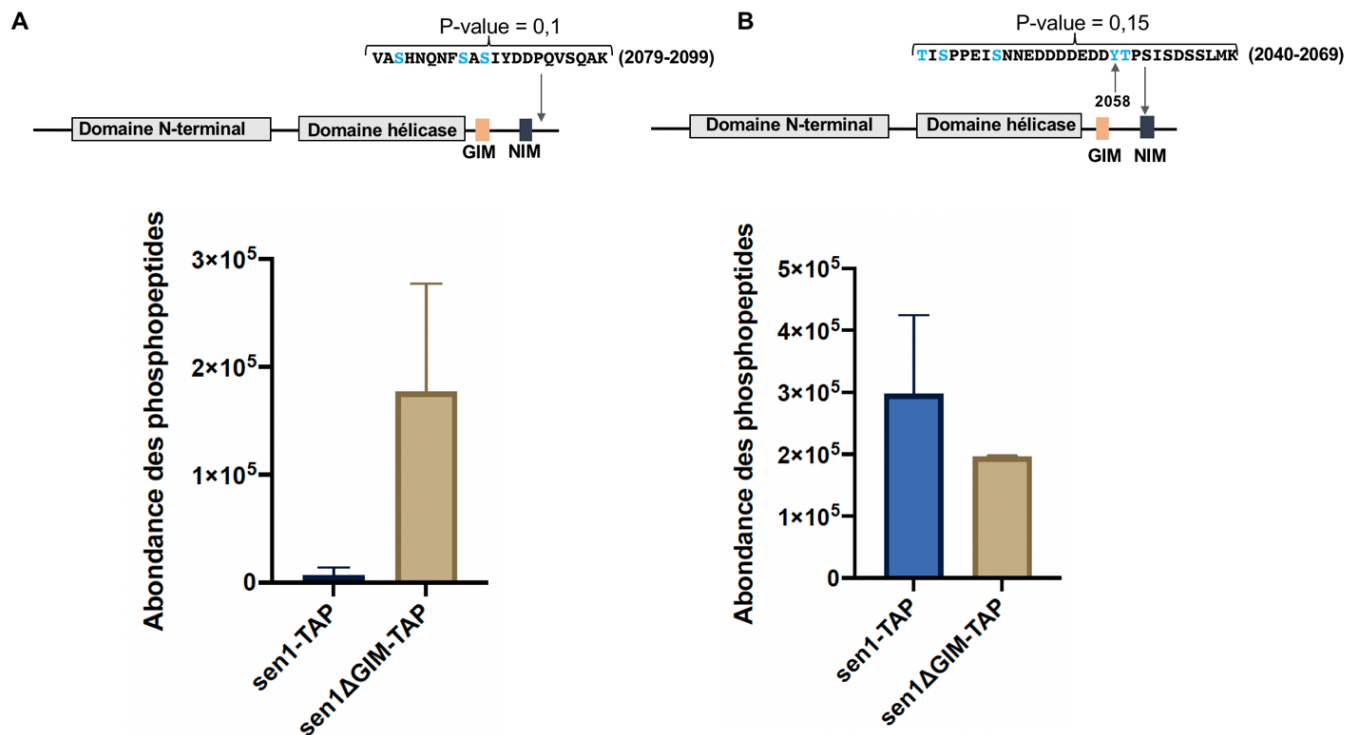


**Figure 36** : La phosphatase Glc7 module l'interaction de Sen1 avec Nrd1 et Nab3.

A : Volcano plot représentant les partenaires majeurs de Sen1 détectés dans cette expérience. Le contrôle est la souche non-tagguée. B : Volcano plot illustrant les interactions affectées par la délétion du GIM. Les points verts correspondent aux protéines ayant un p-value inférieur à 0,01 et un « fold change » (FC) supérieur à 2 tandis que les points rouges ont un p-value compris entre 0,01 et 0,05 et un FC supérieur à 2. Quelques partenaires pertinents de Sen1 sont choisis arbitrairement puis identifiés sur le volcano plot.

Nous avons par la suite analysé l'état de phosphorylation de Sen1 dans ce mutant comparé à la souche sauvage. Cette analyse a révélé l'augmentation d'une phosphorylation dans la région C-terminale de Sen1 dans le mutant (figure 37A) qui pourrait être la cible de Glc7. Ces résidus se trouvent à proximité du GIM et aussi du NIM et donc, cette phosphorylation pourrait participer à la modulation de l'interaction de Sen1 avec Nrd1 et Nab3. Nous avons observé aussi une légère diminution du niveau de phosphorylation de Sen1 au niveau des résidus localisés dans la région du NIM sans préciser la position du site de phosphorylation (figure 37B). Néanmoins, parmi les résidus potentiellement phosphorylés figurent la tyrosine Y2058, que nous avons proposé comme cible putative de Glc7. Ce résultat suggère ainsi que ce dernier n'est pas déphosphorylé par Glc7, contrairement à notre hypothèse de départ. Comme indiqué précédemment, il a été proposé que la phosphorylation de la tyrosine 1 du CTD de l'ARNpol II diminue le recrutement de Nrd1 (Mayer et al., 2012). Cette diminution du niveau de phosphorylation de la Y en position 2058 du NIM pourrait expliquer l'augmentation de l'interaction de Sen1 avec Nrd1 et Nab3.

L'ensemble de ces données nous a conduit à supposer que Glc7 pourrait participer à la régulation de l'interaction de Sen1 avec Nrd1 et Nab3 soit par la déphosphorylation des résidus dans la région 2079-2099, soit par encombrement stérique, c'est-à-dire que la présence de Glc7 empêcherait la liaison de Nrd1 et Nab3. Il est important de mentionner que les valeurs des p-values étant élevées, il est difficile de conclure sur la participation de ces phosphorylations dans la régulation de cette interaction. Ces expériences devraient être répétées pour avoir de résultats plus clairs. Néanmoins, ces données nous donnent des pistes sur le rôle possible de Glc7 dans la régulation de l'interaction de Sen1 avec Nrd1 et Nab3.



**Figure 37** : Analyse de l'état de phosphorylation de Sen1 et *sen1ΔGIM*.

En haut la séquence de la protéine Sen1 avec les phosphopeptides identifiés en spectrométrie de masse avec les valeurs des p-values. Les phosphopeptides dans chaque échantillon ont été identifiés puis quantifiés. La position du site de phosphorylation n'a pas été clairement identifié dans cette expérience. Ainsi, les résidus marqués en bleu sont celles qui sont potentiellement phosphorylés. Les histogrammes (A et B) représentent la quantité de chaque phosphopeptide identifiés en comparant *sen1-TAP* à *sen1ΔGIM-TAP*.

### 3.3.5 Résumé et perspectives :

Dans cette étude, nous avons tenté de caractériser le rôle de l'interaction de Sen1 avec un de ses partenaires majeurs, la phosphatase Glc7. Tout d'abord, nous avons délété le motif GIM et l'analysé l'effet de cette mutation combinée à la délétion de *RRP6* sur la croissance, qui a révélé une très légère augmentation du défaut de croissance du double mutant par rapport à  $\Delta rrp6$  à 37°C. Ce résultat suggère une participation possible de cette interaction à la fonction de la voie NNS.

Nos analyses transcriptomiques ont révélé des défauts de terminaison très modestes pour certaines cibles de la voie NNS. Le nombre d'ARN non-codants qui présentent des défauts étant trop petit, nous ne sommes pas en mesure de réaliser des analyses bio-informatiques plus poussées afin de déterminer les caractéristiques communes de ces cibles et comprendre leur sensibilité au manque d'interaction entre Sen1 et Glc7.

Enfin, par des approches protéomiques, nous avons révélé que la phosphatase Glc7 pourrait moduler l'interaction de Sen1 avec Nrd1 et Nab3 mais des questions sur les mécanismes de cette modulation restent ouvertes. D'après le modèle sur la terminaison de la transcription non-codante, Nrd1 et Nab3 reconnaissent des motifs spécifiques sur l'ARN naissant, ce qui favoriserait le recrutement de Sen1 qui transloque l'ARN et dissocie le complexe d'élongation afin de terminer la transcription. Cependant, cela va à l'encontre des observations dans notre article récente qui montrent que l'interaction de Sen1 avec Nrd1 et Nab3 n'est pas essentielle pour le recrutement de Sen1 (Han et al., 2020). Nrd1 et Nab3 resteraient associés à l'ARN naissant pour promouvoir le recrutement des

enzymes de dégradation de l'ARN. Cependant la question sur les mécanismes qui permettent la dissociation de Sen1 de Nrd1 et Nab3 reste ouverte. Nous avons montré que la perte d'interaction de Sen1 avec Glc7 est aussi corrélée à une légère augmentation de l'association de Sen1 avec Nrd1 et Nab3. Nos données bien que préliminaires nous laissent imaginer que Glc7 pourrait participer au relâchement de Sen1 de Nrd1 et Nab3 pour un certain nombre d'ARN non-codants et de ce fait, contribuer à rendre plus efficace la terminaison de la transcription non-codante. Cependant, le fait que la grande majorité des gènes non-codants ne montrent pas de défaut de terminaison dans le mutant *sen1ΔGIM* suggère que le rôle de Glc7 pourrait être redondant avec d'autres mécanismes et/ou que cette fonction serait nécessaire uniquement dans certaines conditions physiologiques qui n'ont pas été analysées dans cette étude.

Par ailleurs, afin de mieux comprendre cette fonction possible de Glc7, il serait intéressant d'introduire des mutations dans les sites identifiés comme des cibles putatives de Glc7 pour mimer la phosphorylation et analyser l'effet de ces mutations dans l'interaction de Sen1 avec Nrd1 et Nab3. Ces expériences devraient pouvoir nous indiquer si l'effet de la délétion du GIM sur l'interaction de Sen1 avec ses partenaires est dû à l'augmentation des niveaux de phosphorylation dans les résidus identifiés. Si le mutant phosphomimétique ne présentait pas une différence dans les niveaux d'association avec Nrd1 et Nab3, l'hypothèse de l'encombrement stérique serait privilégiée. De façon complémentaire, il pourrait être testé avec de souches qui permettent de dépléter la protéine Glc7 et qui sont disponibles dans notre laboratoire si la déplétion de Glc7 provoque une augmentation de l'interaction de Sen1 avec Nrd1 et Nab3.

Pour conclure, il reste possible que l'interaction de Sen1 avec Glc7 soit nécessaire pour d'autres fonctions de Sen1 et non pas pour son rôle dans la terminaison de la transcription non-codante. En effet, il a été montré que Sen1 joue un rôle dans la résolution des R-loops, des conflits entre la réplication et la transcription et dans la réparation de l'ADN (Appanah et al., 2020; Mischo et al., 2011; Rawal et al., 2020). Il serait donc intéressant de faire un crible génétique afin d'analyser s'il y'a des interactions génétiques entre le mutant *sen1ΔGIM* et les mutants affectés dans ces différentes fonctions dans le but d'explorer un possible rôle de cette interaction dans ces processus.

## 4 Bibliographie



## Bibliographie :

- Akhtar, M.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D., Ansari, A.Z., 2009. TFIIF kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol Cell* 34, 387–393. <https://doi.org/10.1016/j.molcel.2009.04.016>
- Alpert, T., Straube, K., Carrillo Oesterreich, F., Herzel, L., Neugebauer, K.M., 2020. Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns. *Cell Reports* 33, 108324. <https://doi.org/10.1016/j.celrep.2020.108324>
- Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. <https://doi.org/10.1093/bioinformatics/btu638>
- André, K.M., Sipos, E.H., Soutourina, J., 2020. Mediator Roles Going Beyond Transcription. *Trends Genet.* <https://doi.org/10.1016/j.tig.2020.08.015>
- Appanah, R., Lones, E.C., Aiello, U., Libri, D., De Piccoli, G., 2020. Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability. *Cell Reports* 30, 2094–2105.e9. <https://doi.org/10.1016/j.celrep.2020.01.087>
- Arigo, J.T., Carroll, K.L., Ames, J.M., Corden, J.L., 2006a. Regulation of Yeast NRD1 Expression by Premature Transcription Termination. *Molecular Cell* 21, 641–651. <https://doi.org/10.1016/j.molcel.2006.02.005>
- Arigo, J.T., Eyler, D.E., Carroll, K.L., Corden, J.L., 2006b. Termination of Cryptic Unstable Transcripts Is Directed by Yeast RNA-Binding Proteins Nrd1 and Nab3. *Molecular Cell* 23, 841–851. <https://doi.org/10.1016/j.molcel.2006.07.024>
- Arndt, K.M., Reines, D., 2015a. Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annu Rev Biochem* 84, 381–404. <https://doi.org/10.1146/annurev-biochem-060614-034457>
- Arndt, K.M., Reines, D., 2015b. Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annual Review of Biochemistry* 84, 381–404. <https://doi.org/10.1146/annurev-biochem-060614-034457>
- Assenholt, J., Mouaikel, J., Andersen, K.R., Brodersen, D.E., Libri, D., Jensen, T.H., 2008. Exonucleolysis is required for nuclear mRNA quality control in yeast THO mutants. *RNA* 14, 2305–2313. <https://doi.org/10.1261/rna.1108008>
- Bacikova, V., Pasulka, J., Kubicek, K., Stefl, R., 2014. Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res* 42, 8024–8038. <https://doi.org/10.1093/nar/gku446>
- Bae, H.J., Dubarry, M., Jeon, J., Soares, L.M., Dargemont, C., Kim, J., Geli, V., Buratowski, S., 2020. The Set1 N-terminal domain and Swd2 interact with RNA polymerase II CTD to recruit COMPASS. *Nat Commun* 11, 2181. <https://doi.org/10.1038/s41467-020-16082-2>
- Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S., Söding, J., Cramer, P., 2017. Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol Cell* 66, 38–49.e6. <https://doi.org/10.1016/j.molcel.2017.02.009>

- Baejen, C., Torkler, P., Gressel, S., Essig, K., Söding, J., Cramer, P., 2014. Transcriptome Maps of mRNP Biogenesis Factors Define Pre-mRNA Recognition. *Molecular Cell* 55, 745–757. <https://doi.org/10.1016/j.molcel.2014.08.005>
- Baptista, T., Grünberg, S., Minoungou, N., Koster, M.J.E., Timmers, H.T.M., Hahn, S., Devys, D., Tora, L., 2017. SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Molecular Cell* 68, 130-143.e5. <https://doi.org/10.1016/j.molcel.2017.08.016>
- Basehoar, A.D., Zanton, S.J., Pugh, B.F., 2004. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. *Cell* 116, 699–709. [https://doi.org/10.1016/S0092-8674\(04\)00205-3](https://doi.org/10.1016/S0092-8674(04)00205-3)
- Bataille, A.R., Jeronimo, C., Jacques, P.-É., Laramée, L., Fortin, M.-È., Forest, A., Bergeron, M., Hanes, S.D., 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. <https://doi.org/10.1016/j.molcel.2011.11.024>
- Bennett, C.L., La Spada, A.R., 2018. Senataxin, A Novel Helicase at the Interface of RNA Transcriptome Regulation and Neurobiology: From Normal Function to Pathological Roles in Motor Neuron Disease and Cerebellar Degeneration. *Adv Neurobiol* 20, 265–281. [https://doi.org/10.1007/978-3-319-89689-2\\_10](https://doi.org/10.1007/978-3-319-89689-2_10)
- Benoist, C., Chambon, P., 1981. In vivo sequence requirements of the SV40 early promoter region. *Nature* 290, 304–310. <https://doi.org/10.1038/290304a0>
- Berretta, J., Pinskaya, M., Morillon, A., 2008. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes & Development* 22, 615–626. <https://doi.org/10.1101/gad.458008>
- Bird, Amanda J, Gordon, M., Eide, D.J., Winge, D.R., 2006. Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. *EMBO J* 25, 5726–5734. <https://doi.org/10.1038/sj.emboj.7601453>
- Bird, Amanda J., Gordon, M., Eide, D.J., Winge, D.R., 2006. Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. *EMBO J* 25, 5726–5734. <https://doi.org/10.1038/sj.emboj.7601453>
- Boeger, H., Griesenbeck, J., Kornberg, R.D., 2008. Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell* 133, 716–726. <https://doi.org/10.1016/j.cell.2008.02.051>
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L., Devys, D., 2014. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev* 28, 1999–2012. <https://doi.org/10.1101/gad.250225.114>
- Bresson, S., Tuck, A., Staneva, D., Tollervy, D., 2017. Nuclear RNA Decay Pathways Aid Rapid Remodeling of Gene Expression in Yeast. *Mol. Cell* 65, 787-800.e5. <https://doi.org/10.1016/j.molcel.2017.01.005>

- Briggs, S.D., 2001. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes & Development* 15, 3286–3295. <https://doi.org/10.1101/gad.940201>
- Bumgarner, S.L., Dowell, R.D., Grisafi, P., Gifford, D.K., Fink, G.R., 2009. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proceedings of the National Academy of Sciences* 106, 18321–18326. <https://doi.org/10.1073/pnas.0909641106>
- Burke, T.W., Kadonaga, J.T., 1997. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes & Development* 11, 3020–3031. <https://doi.org/10.1101/gad.11.22.3020>
- Cabarcas, S., Schramm, L., 2011. RNA polymerase III transcription in cancer: the BRF2 connection. *Mol Cancer* 10, 47. <https://doi.org/10.1186/1476-4598-10-47>
- Cabart, P., Ujvari, A., Pal, M., Luse, D.S., 2011. 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. *Proceedings of the National Academy of Sciences* 108, 15786–15791. <https://doi.org/10.1073/pnas.1104591108>
- Cai, G., Imasaki, T., Yamada, K., Cardelli, F., Takagi, Y., Asturias, F.J., 2010. Mediator head module structure and functional interactions. *Nat Struct Mol Biol* 17, 273–279. <https://doi.org/10.1038/nsmb.1757>
- Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., Stutz, F., 2007a. Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in *S. cerevisiae*. *Cell* 131, 706–717. <https://doi.org/10.1016/j.cell.2007.09.014>
- Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., Stutz, F., 2007b. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* 131, 706–717. <https://doi.org/10.1016/j.cell.2007.09.014>
- Candelli, T., Challal, D., Briand, J.-B., Boulay, J., Porrua, O., Colin, J., Libri, D., 2018a. High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. *EMBO J.* 37. <https://doi.org/10.15252/embj.201797490>
- Candelli, T., Gros, J., Libri, D., 2018b. Pervasive transcription fine-tunes replication origin activity. *Elife* 7. <https://doi.org/10.7554/eLife.40802>
- Carey, M., Li, B., Workman, J.L., 2006. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Mol Cell* 24, 481–487. <https://doi.org/10.1016/j.molcel.2006.09.012>
- Carroll, K.L., Ghirlando, R., Ames, J.M., Corden, J.L., 2007. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* 13, 361–373. <https://doi.org/10.1261/rna.338407>
- Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F., Zenklusen, D., 2013a. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* 20, 851–858. <https://doi.org/10.1038/nsmb.2598>
- Castelnuovo, M., Rahman, S., Guffanti, E., Infantino, V., Stutz, F., Zenklusen, D., 2013b. Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat. Struct. Mol.*

- Biol. 20, 851–858. <https://doi.org/10.1038/nsmb.2598>
- Castelnuovo, M., Zaugg, J.B., Guffanti, E., Maffioletti, A., Camblong, J., Xu, Z., Clauder-Münster, S., Steinmetz, L.M., Luscombe, N.M., Stutz, F., 2014a. Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Research* 42, 4348–4362. <https://doi.org/10.1093/nar/gku100>
- Castelnuovo, M., Zaugg, J.B., Guffanti, E., Maffioletti, A., Camblong, J., Xu, Z., Clauder-Münster, S., Steinmetz, L.M., Luscombe, N.M., Stutz, F., 2014b. Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic Acids Res.* 42, 4348–4362. <https://doi.org/10.1093/nar/gku100>
- Cavallini, B., Huet, J., Plassat, J.-L., Sentenac, A., Egly, J.-M., Chambon, P., 1988. A yeast activity can substitute for the HeLa cell TATA box factor. *Nature* 334, 77–80. <https://doi.org/10.1038/334077a0>
- Challal, D., Barucco, M., Kubik, S., Feuerbach, F., Candelli, T., Geoffroy, H., Benaksas, C., Shore, D., Libri, D., 2018. General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. *Mol. Cell* 72, 955-969.e7. <https://doi.org/10.1016/j.molcel.2018.11.037>
- Chang, Y.-W., Howard, S.C., Herman, P.K., 2004. The Ras/PKA signaling pathway directly targets the Srb9 protein, a component of the general RNA polymerase II transcription apparatus. *Mol Cell* 15, 107–116. <https://doi.org/10.1016/j.molcel.2004.05.021>
- Chen, H.-T., Hahn, S., 2004. Mapping the Location of TFIIB within the RNA Polymerase II Transcription Preinitiation Complex. *Cell* 119, 169–180. <https://doi.org/10.1016/j.cell.2004.09.028>
- Chen, X., Müller, U., Sundling, K.E., Brow, D.A., 2014. *Saccharomyces cerevisiae* Sen1 as a Model for the Study of Mutations in Human Senataxin That Elicit Cerebellar Ataxia. *Genetics* 198, 577–590. <https://doi.org/10.1534/genetics.114.167585>
- Cheon, Y., Kim, H., Park, K., Kim, M., Lee, D., 2020. Dynamic modules of the coactivator SAGA in eukaryotic transcription. *Exp Mol Med* 52, 991–1003. <https://doi.org/10.1038/s12276-020-0463-4>
- Cheung, A.C.M., Cramer, P., 2011. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* 471, 249–253. <https://doi.org/10.1038/nature09785>
- Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., Culbertson, M.R., 2012. Interactions of Sen1, Nrd1, and Nab3 with Multiple Phosphorylated Forms of the Rpb1 C-Terminal Domain in *Saccharomyces cerevisiae*. *Eukaryotic Cell* 11, 417–429. <https://doi.org/10.1128/EC.05320-11>
- Chong, Y.T., Koh, J.L.Y., Friesen, H., Kaluarachchi Duffy, S., Cox, M.J., Moses, A., Moffat, J., Boone, C., Andrews, B.J., 2015. Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. *Cell* 161, 1413–1424. <https://doi.org/10.1016/j.cell.2015.04.051>
- Chow, C.-W., Davis, R.J., 2006. Proteins kinases: chromatin-associated enzymes? *Cell* 127, 887–890. <https://doi.org/10.1016/j.cell.2006.11.015>
- Churchman, L.S., Weissman, J.S., 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–373. <https://doi.org/10.1038/nature09652>

- Clotet, J., Posas, F., 2007. Control of Cell Cycle in Response to Osmostress: Lessons from Yeast, in: *Methods in Enzymology*. Elsevier, pp. 63–76. [https://doi.org/10.1016/S0076-6879\(07\)28004-8](https://doi.org/10.1016/S0076-6879(07)28004-8)
- Colin, J., Libri, D., Porrua, O., 2011. Cryptic transcription and early termination in the control of gene expression. *Genet Res Int* 2011, 653494. <https://doi.org/10.4061/2011/653494>
- Collin, P., Jeronimo, C., Poitras, C., Robert, F., 2019. RNA Polymerase II CTD Tyrosine 1 Is Required for Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. *Mol Cell* 73, 655-669.e7. <https://doi.org/10.1016/j.molcel.2018.12.002>
- Conaway, R.C., Garrett, K.P., Hanley, J.P., Conaway, J.W., 1991. Mechanism of promoter selection by RNA polymerase II: mammalian transcription factors alpha and beta gamma promote entry of polymerase into the preinitiation complex. *Proceedings of the National Academy of Sciences* 88, 6205–6209. <https://doi.org/10.1073/pnas.88.14.6205>
- Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., Corden, J.L., 2000. A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* 154, 557–571.
- Corden, J.L., 2013. RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template. *Chem. Rev.* 113, 8423–8455. <https://doi.org/10.1021/cr400158h>
- Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., 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. <https://doi.org/10.1371/journal.pgen.1002329>
- Darby, M. M., Serebreni, L., Pan, X., Boeke, J.D., Corden, J.L., 2012. The *Saccharomyces cerevisiae* Nrd1-Nab3 Transcription Termination Pathway Acts in Opposition to Ras Signaling and Mediates Response to Nutrient Depletion. *Molecular and Cellular Biology* 32, 1762–1775. <https://doi.org/10.1128/MCB.00050-12>
- Darby, Miranda M., Serebreni, L., Pan, X., Boeke, J.D., Corden, J.L., 2012. The *Saccharomyces cerevisiae* Nrd1-Nab3 transcription termination pathway acts in opposition to Ras signaling and mediates response to nutrient depletion. *Mol. Cell. Biol.* 32, 1762–1775. <https://doi.org/10.1128/MCB.00050-12>
- David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., Steinmetz, L.M., 2006. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A* 103, 5320–5325. <https://doi.org/10.1073/pnas.0601091103>
- de Nadal, E., Zapater, M., Alepuz, P.M., Sumoy, L., Mas, G., Posas, F., 2004. The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoresponsive genes. *Nature* 427, 370–374. <https://doi.org/10.1038/nature02258>
- Dechassa, M.L., Sabri, A., Pondugula, S., Kassabov, S.R., Chatterjee, N., Kladdé, M.P., Bartholomew, B., 2010. SWI/SNF Has Intrinsic Nucleosome Disassembly Activity that Is Dependent on Adjacent Nucleosomes. *Molecular Cell* 38, 590–602. <https://doi.org/10.1016/j.molcel.2010.02.040>
- Delan-Forino, C., Spanos, C., Rappsilber, J., Tollervy, D., 2020. Substrate specificity of the TRAMP nuclear surveillance complexes. *Nat Commun* 11, 3122. <https://doi.org/10.1038/s41467-020-16965-4>

Deng, W., 2005. A core promoter element downstream of the TATA box that is recognized by TFIIB. *Genes & Development* 19, 2418–2423. <https://doi.org/10.1101/gad.342405>

Dengl, S., Cramer, P., 2009. Torpedo nuclease Rat1 is insufficient to terminate RNA polymerase II in vitro. *J Biol Chem* 284, 21270–21279. <https://doi.org/10.1074/jbc.M109.013847>

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakraborty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigó, R., Gingeras, T.R., 2012. Landscape of transcription in human cells. *Nature* 489, 101–108. <https://doi.org/10.1038/nature11233>

Dosztanyi, Z., Csizmok, V., Tompa, P., Simon, I., 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21, 3433–3434. <https://doi.org/10.1093/bioinformatics/bti541>

Dosztányi, Z., Csizmok, V., Tompa, P., Simon, I., 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21, 3433–3434. <https://doi.org/10.1093/bioinformatics/bti541>

du Mee, D.J.M., Ivanov, M., Parker, J.P., Buratowski, S., Marquardt, S., 2018. Efficient termination of nuclear lncRNA transcription promotes mitochondrial genome maintenance. *eLife* 7, e31989. <https://doi.org/10.7554/eLife.31989>

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., West, S., 2018. Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. *Genes Dev.* 32, 127–139. <https://doi.org/10.1101/gad.308528.117>

Egloff, S., Dienstbier, M., Murphy, S., 2012. Updating the RNA polymerase CTD code: adding gene-specific layers. *Trends in Genetics* 28, 333–341. <https://doi.org/10.1016/j.tig.2012.03.007>

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., Werner, M., 2008. Mediator-dependent recruitment of TFIIB modules in preinitiation complex. *Mol Cell* 31, 337–346. <https://doi.org/10.1016/j.molcel.2008.06.021>

Eyboulet, F., Wydau-Dematteis, S., Eychenne, T., Alibert, O., Neil, H., Boschiero, C., Nevers, M.-C., Volland, H., Cornu, D., Redeker, V., Werner, M., Soutourina, J., 2015. Mediator independently orchestrates multiple steps of preinitiation complex assembly in vivo. *Nucleic Acids Res* 43, 9214–9231. <https://doi.org/10.1093/nar/gkv782>

Eychenne, T., Novikova, E., Barrault, M.-B., Alibert, O., Boschiero, C., Peixeiro, N., Cornu, D., Redeker, V., Kuras, L., Nicolas, P., Werner, M., Soutourina, J., 2016. Functional interplay between Mediator and TFIIB in preinitiation complex assembly in relation to promoter architecture. *Genes Dev* 30, 2119–2132. <https://doi.org/10.1101/gad.285775.116>

- Fairman-Williams, M.E., Guenther, U.-P., Jankowsky, E., 2010. SF1 and SF2 helicases: family matters. *Current Opinion in Structural Biology* 20, 313–324. <https://doi.org/10.1016/j.sbi.2010.03.011>
- Falk, S., Weir, J.R., Hentschel, J., Reichelt, P., Bonneau, F., Conti, E., 2014. The molecular architecture of the TRAMP complex reveals the organization and interplay of its two catalytic activities. *Mol Cell* 55, 856–867. <https://doi.org/10.1016/j.molcel.2014.07.020>
- Fasken, M.B., Laribee, R.N., Corbett, A.H., 2015. Nab3 Facilitates the Function of the TRAMP Complex in RNA Processing via Recruitment of Rrp6 Independent of Nrd1. *PLoS Genet* 11, e1005044. <https://doi.org/10.1371/journal.pgen.1005044>
- Fernandes, J.C.R., Acuña, S.M., Aoki, J.I., Floeter-Winter, L.M., Muxel, S.M., 2019. Long Non-Coding RNAs in the Regulation of Gene Expression: Physiology and Disease. *Noncoding RNA* 5. <https://doi.org/10.3390/ncrna5010017>
- Finkel, J.S., Chinchilla, K., Ursic, D., Culbertson, M.R., 2010. Sen1p Performs Two Genetically Separable Functions in Transcription and Processing of U5 Small Nuclear RNA in *Saccharomyces cerevisiae*. *Genetics* 184, 107–118. <https://doi.org/10.1534/genetics.109.110031>
- Finoux, A.-L., Séraphin, B., 2006. In vivo targeting of the yeast Pop2 deadenylase subunit to reporter transcripts induces their rapid degradation and generates new decay intermediates. *J. Biol. Chem.* 281, 25940–25947. <https://doi.org/10.1074/jbc.M600132200>
- Fleming, A.B., Kao, C.-F., Hillyer, C., Pikaart, M., Osley, M.A., 2008. H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Mol Cell* 31, 57–66. <https://doi.org/10.1016/j.molcel.2008.04.025>
- Fong, N., Brannan, K., Erickson, B., Kim, H., Cortazar, M.A., Sheridan, R.M., Nguyen, T., Karp, S., Bentley, D.L., 2015. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Mol Cell* 60, 256–267. <https://doi.org/10.1016/j.molcel.2015.09.026>
- Frey, A.G., Eide, D.J., 2012. Zinc-responsive coactivator recruitment by the yeast Zap1 transcription factor. *Microbiologyopen* 1, 105–114. <https://doi.org/10.1002/mbo3.8>
- Gary, J.D., Clarke, S., 1998. RNA and Protein Interactions Modulated by Protein Arginine Methylation, in: *Progress in Nucleic Acid Research and Molecular Biology*. Elsevier, pp. 65–131. [https://doi.org/10.1016/S0079-6603\(08\)60825-9](https://doi.org/10.1016/S0079-6603(08)60825-9)
- Gelfand, B., Mead, J., Bruning, A., Apostolopoulos, N., Tadigotla, V., Nagaraj, V., Sengupta, A.M., Vershon, A.K., 2011. Regulated Antisense Transcription Controls Expression of Cell-Type-Specific Genes in Yeast. *Molecular and Cellular Biology* 31, 1701–1709. <https://doi.org/10.1128/MCB.01071-10>
- Gershenzon, N.I., Ioshikhes, I.P., 2005. Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics* 21, 1295–1300. <https://doi.org/10.1093/bioinformatics/bti172>
- Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O’Shea, E.K., Weissman, J.S., 2003. Global analysis of protein expression in yeast. *Nature* 425, 737–741. <https://doi.org/10.1038/nature02046>

- Ghazy, M.A., Brodie, S.A., Ammerman, M.L., Ziegler, L.M., Ponticelli, A.S., 2004. Amino Acid Substitutions in Yeast TFIIF Confer Upstream Shifts in Transcription Initiation and Altered Interaction with RNA Polymerase II. *MCB* 24, 10975–10985. <https://doi.org/10.1128/MCB.24.24.10975-10985.2004>
- Gill, J.K., Maffioletti, A., García-Molinero, V., Stutz, F., Soudet, J., 2020. Fine Chromatin-Driven Mechanism of Transcription Interference by Antisense Noncoding Transcription. *Cell Reports* 31, 107612. <https://doi.org/10.1016/j.celrep.2020.107612>
- Ginsburg, D.S., Govind, C.K., Hinnebusch, A.G., 2009. NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5. *Mol Cell Biol* 29, 6473–6487. <https://doi.org/10.1128/MCB.01033-09>
- Goodfellow, S.J., Zomerdijk, J.C.B.M., 2013. Basic Mechanisms in RNA Polymerase I Transcription of the Ribosomal RNA Genes, in: Kundu, T.K. (Ed.), *Epigenetics: Development and Disease, Subcellular Biochemistry*. Springer Netherlands, Dordrecht, pp. 211–236. [https://doi.org/10.1007/978-94-007-4525-4\\_10](https://doi.org/10.1007/978-94-007-4525-4_10)
- Govind, C.K., Zhang, F., Qiu, H., Hofmeyer, K., Hinnebusch, A.G., 2007. Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. *Mol Cell* 25, 31–42. <https://doi.org/10.1016/j.molcel.2006.11.020>
- Granneman, S., Kudla, G., Petfalski, E., Tollervey, D., 2009. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9613–9618. <https://doi.org/10.1073/pnas.0901997106>
- Green, D.M., Marfatia, K.A., Crafton, E.B., Zhang, X., Cheng, X., Corbett, A.H., 2002. Nab2p is required for poly(A) RNA export in *Saccharomyces cerevisiae* and is regulated by arginine methylation via Hmt1p. *J Biol Chem* 277, 7752–7760. <https://doi.org/10.1074/jbc.M110053200>
- Groh, M., Albulescu, L.O., Cristini, A., Gromak, N., 2017. Senataxin: Genome Guardian at the Interface of Transcription and Neurodegeneration. *J Mol Biol* 429, 3181–3195. <https://doi.org/10.1016/j.jmb.2016.10.021>
- Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., Libri, D., 2012. Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol Cell* 48, 409–421. <https://doi.org/10.1016/j.molcel.2012.08.018>
- Gupta, K., Sari-Ak, D., Haffke, M., Trowitzsch, S., Berger, I., 2016. Zooming in on Transcription Preinitiation. *J Mol Biol* 428, 2581–2591. <https://doi.org/10.1016/j.jmb.2016.04.003>
- Gurard-Levin, Z.A., Quivy, J.-P., Almouzni, G., 2014. Histone Chaperones: Assisting Histone Traffic and Nucleosome Dynamics. *Annu. Rev. Biochem.* 83, 487–517. <https://doi.org/10.1146/annurev-biochem-060713-035536>
- Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* 11, 394–403. <https://doi.org/10.1038/nsmb763>
- Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M., Martens, J.A., 2011. Intergenic transcription causes repression by directing nucleosome assembly. *Genes & Development* 25, 29–40. <https://doi.org/10.1101/gad.1975011>
- Hamill, S., Wolin, S.L., Reinisch, K.M., 2010. Structure and function of the polymerase core of



TRAMP, a RNA surveillance complex. *Proceedings of the National Academy of Sciences* 107, 15045–15050. <https://doi.org/10.1073/pnas.1003505107>

Han, Z., Jasnovidova, O., Haidara, N., Tudek, A., Kubicek, K., Libri, D., Stefl, R., Porrua, O., 2020. Termination of non-coding transcription in yeast relies on both an RNA Pol II CTD interaction domain and a CTD-mimicking region in Sen1. *EMBO J* 39. <https://doi.org/10.15252/embj.2019101548>

Han, Z., Libri, D., Porrua, O., 2017. Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Research* 45, 1355–1370. <https://doi.org/10.1093/nar/gkw1230>

Hantsche, M., Cramer, P., 2017. Conserved RNA polymerase II initiation complex structure. *Current Opinion in Structural Biology* 47, 17–22. <https://doi.org/10.1016/j.sbi.2017.03.013>

Harlen, K.M., 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. <https://doi.org/10.1038/nrm.2017.10>

Harlen, K.M., Trotta, K.L., Smith, E.E., Mosaheb, M.M., Fuchs, S.M., Churchman, L.S., 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* 15, 2147–2158. <https://doi.org/10.1016/j.celrep.2016.05.010>

Hassan, A.H., Neely, K.E., Workman, J.L., 2001. Histone Acetyltransferase Complexes Stabilize SWI/SNF Binding to Promoter Nucleosomes. *Cell* 104, 817–827. [https://doi.org/10.1016/S0092-8674\(01\)00279-3](https://doi.org/10.1016/S0092-8674(01)00279-3)

Hector, R.E., 2002. Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *The EMBO Journal* 21, 1800–1810. <https://doi.org/10.1093/emboj/21.7.1800>

Heidemann, M., Hintermair, C., Voß, K., Eick, D., 2013. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim Biophys Acta* 1829, 55–62. <https://doi.org/10.1016/j.bbagr.2012.08.013>

Heo, D., Yoo, I., Kong, J., Lidschreiber, M., Mayer, A., Choi, B.-Y., Hahn, Y., Cramer, P., Buratowski, S., Kim, M., 2013. The RNA Polymerase II C-terminal Domain-interacting Domain of Yeast Nrd1 Contributes to the Choice of Termination Pathway and Couples to RNA Processing by the Nuclear Exosome. *J. Biol. Chem.* 288, 36676–36690. <https://doi.org/10.1074/jbc.M113.508267>

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., Stefl, R., 2011. Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J Biol Chem* 286, 3645–3657. <https://doi.org/10.1074/jbc.M110.158774>

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., Young, R.A., 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728. [https://doi.org/10.1016/s0092-8674\(00\)81641-4](https://doi.org/10.1016/s0092-8674(00)81641-4)

Holstege, F.C.P., 1997. Three transitions in the RNA polymerase II transcription complex during initiation. *The EMBO Journal* 16, 7468–7480. <https://doi.org/10.1093/emboj/16.24.7468>

Hongay, C.F., Grisafi, P.L., Galitski, T., Fink, G.R., 2006. Antisense Transcription Controls Cell Fate in *Saccharomyces cerevisiae*. *Cell* 127, 735–745. <https://doi.org/10.1016/j.cell.2006.09.038>

- Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., Vogelauer, M., 2008. A ncRNA Modulates Histone Modification and mRNA Induction in the Yeast GAL Gene Cluster. *Molecular Cell* 32, 685–695. <https://doi.org/10.1016/j.molcel.2008.09.027>
- Houseley, J., Tollervey, D., 2006. Yeast Trf5p is a nuclear poly(A) polymerase. *EMBO Rep* 7, 205–211. <https://doi.org/10.1038/sj.embor.7400612>
- Hsin, J.-P., Manley, J.L., 2012. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & Development* 26, 2119–2137. <https://doi.org/10.1101/gad.200303.112>
- Huang, Y.-C., Chen, H.-T., Teng, S.-C., 2010. Intragenic transcription of a noncoding RNA modulates expression of ASP3 in budding yeast. *RNA* 16, 2085–2093. <https://doi.org/10.1261/rna.2177410>
- Huber, F., Bunina, D., Gupta, I., Khmelinskii, A., Meurer, M., Theer, P., Steinmetz, L.M., Knop, M., 2016. Protein Abundance Control by Non-coding Antisense Transcription. *Cell Rep* 15, 2625–2636. <https://doi.org/10.1016/j.celrep.2016.05.043>
- Huisinga, K.L., Pugh, B.F., 2004. A Genome-Wide Housekeeping Role for TFIID and a Highly Regulated Stress-Related Role for SAGA in *Saccharomyces cerevisiae*. *Molecular Cell* 13, 573–585. [https://doi.org/10.1016/S1097-2765\(04\)00087-5](https://doi.org/10.1016/S1097-2765(04)00087-5)
- Imbalzano, A.N., Zaret, K.S., Kingston, R.E., 1994. Transcription factor (TF) IIB and TFIIA can independently increase the affinity of the TATA-binding protein for DNA. *J. Biol. Chem.* 269, 8280–8286.
- Jackson, R.N., Klauer, A.A., Hintze, B.J., Robinson, H., van Hoof, A., Johnson, S.J., 2010. The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J* 29, 2205–2216. <https://doi.org/10.1038/emboj.2010.107>
- Jensen, L.T., Ajua-Alemanji, M., Culotta, V.C., 2003. The *Saccharomyces cerevisiae* high affinity phosphate transporter encoded by PHO84 also functions in manganese homeostasis. *J Biol Chem* 278, 42036–42040. <https://doi.org/10.1074/jbc.M307413200>
- Jensen, Torben Heick, Jacquier, A., Libri, D., 2013. Dealing with Pervasive Transcription. *Molecular Cell* 52, 473–484. <https://doi.org/10.1016/j.molcel.2013.10.032>
- Jensen, Torben Heick, Jacquier, A., Libri, D., 2013. Dealing with pervasive transcription. *Mol. Cell* 52, 473–484. <https://doi.org/10.1016/j.molcel.2013.10.032>
- Jeronimo, C., Bataille, A.R., Robert, F., 2013. The Writers, Readers, and Functions of the RNA Polymerase II C-Terminal Domain Code. *Chem. Rev.* 113, 8491–8522. <https://doi.org/10.1021/cr4001397>
- Jeronimo, C., Robert, F., 2014. Kin28 regulates the transient association of Mediator with core promoters. *Nat Struct Mol Biol* 21, 449–455. <https://doi.org/10.1038/nsmb.2810>
- Jessen, W.J., Hoose, S.A., Kilgore, J.A., Kladde, M.P., 2006. Active PHO5 chromatin encompasses variable numbers of nucleosomes at individual promoters. *Nat Struct Mol Biol* 13, 256–263. <https://doi.org/10.1038/nsmb1062>
- Juven-Gershon, T., Kadonaga, J.T., 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* 339, 225–229.

<https://doi.org/10.1016/j.ydbio.2009.08.009>

Kaplan, C.D., Laprade, L., Winston, F., 2003. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301, 1096–1099. <https://doi.org/10.1126/science.1087374>

Kasten, M., Szerlong, H., Erdjument-Bromage, H., Tempst, P., Werner, M., Cairns, B.R., 2004. Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. *EMBO J* 23, 1348–1359. <https://doi.org/10.1038/sj.emboj.7600143>

Khapersky, D.A., Ammerman, M.L., Majovski, R.C., Ponticelli, A.S., 2008. Functions of *Saccharomyces cerevisiae* TFIIF during Transcription Start Site Utilization. *MCB* 28, 3757–3766. <https://doi.org/10.1128/MCB.02272-07>

Kilchert, C., Wittmann, S., Vasiljeva, L., 2016. The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol* 17, 227–239. <https://doi.org/10.1038/nrm.2015.15>

Kim, H.J., Khalimonchuk, O., Smith, P.M., Winge, D.R., 2012. Structure, function, and assembly of heme centers in mitochondrial respiratory complexes. *Biochim Biophys Acta* 1823, 1604–1616. <https://doi.org/10.1016/j.bbamcr.2012.04.008>

Kim, K.-Y., Levin, D.E., 2011. Mpk1 MAPK Association with the Paf1 Complex Blocks Sen1-Mediated Premature Transcription Termination. *Cell* 144, 745–756. <https://doi.org/10.1016/j.cell.2011.01.034>

Kim, M., Ahn, S.-H., Krogan, N.J., Greenblatt, J.F., Buratowski, S., 2004. Transitions in RNA polymerase II elongation complexes at the 3' ends of genes. *EMBO J* 23, 354–364. <https://doi.org/10.1038/sj.emboj.7600053>

Kim, M., Suh, H., Cho, E.-J., Buratowski, S., 2009. Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J Biol Chem* 284, 26421–26426. <https://doi.org/10.1074/jbc.M109.028993>

Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H., Kornberg, R.D., 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599–608. [https://doi.org/10.1016/0092-8674\(94\)90221-6](https://doi.org/10.1016/0092-8674(94)90221-6)

Kireeva, M.L., Lubkowska, L., Komissarova, N., Kashlev, M., 2003. Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from *Saccharomyces cerevisiae*. *Meth. Enzymol.* 370, 138–155. [https://doi.org/10.1016/S0076-6879\(03\)70012-3](https://doi.org/10.1016/S0076-6879(03)70012-3)

Komarnitsky, P., 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes & Development* 14, 2452–2460. <https://doi.org/10.1101/gad.824700>

Kornberg, R.D., 2005. Mediator and the mechanism of transcriptional activation. *Trends Biochem Sci* 30, 235–239. <https://doi.org/10.1016/j.tibs.2005.03.011>

Kornberg, R.D., Thonmas, J.O., 1974. Chromatin Structure: Oligomers of the Histones. *Science* 184, 865–868. <https://doi.org/10.1126/science.184.4139.865>

Kouzarides, T., 2007. Chromatin modifications and their function. *Cell* 128, 693–705. <https://doi.org/10.1016/j.cell.2007.02.005>

- Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., Greenblatt, J., 2003. Methylation of Histone H3 by Set2 in *Saccharomyces cerevisiae* Is Linked to Transcriptional Elongation by RNA Polymerase II. *MCB* 23, 4207–4218. <https://doi.org/10.1128/MCB.23.12.4207-4218.2003>
- Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., Stefl, R., 2012. Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes & Development* 26, 1891–1896. <https://doi.org/10.1101/gad.192781.112>
- Kuehner, J.N., Pearson, E.L., Moore, C., 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nat Rev Mol Cell Biol* 12, 283–294. <https://doi.org/10.1038/nrm3098>
- LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., Tollervey, D., 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* 121, 713–724. <https://doi.org/10.1016/j.cell.2005.04.029>
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., Ebright, R.H., 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes & Development* 12, 34–44. <https://doi.org/10.1101/gad.12.1.34>
- Lee, K.Y., Chopra, A., Burke, G.L., Chen, Z., Greenblatt, J.F., Biggar, K.K., Meneghini, M.D., 2020a. A crucial RNA-binding lysine residue in the Nab3 RRM domain undergoes SET1 and SET3-responsive methylation. *Nucleic Acids Res* 48, 2897–2911. <https://doi.org/10.1093/nar/gkaa029>
- Lee, K.Y., Chopra, A., Burke, G.L., Chen, Z., Greenblatt, J.F., Biggar, K.K., Meneghini, M.D., 2020b. A crucial RNA-binding lysine residue in the Nab3 RRM domain undergoes SET1 and SET3-responsive methylation. *Nucleic Acids Res* 48, 2897–2911. <https://doi.org/10.1093/nar/gkaa029>
- Lee, T.I., Young, R.A., 2000. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* 34, 77–137. <https://doi.org/10.1146/annurev.genet.34.1.77>
- Leonaitė, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O., Conti, E., 2017. Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. *EMBO J* 36, 1590–1604. <https://doi.org/10.15252/embj.201696174>
- Li, B., Carey, M., Workman, J.L., 2007. The Role of Chromatin during Transcription. *Cell* 128, 707–719. <https://doi.org/10.1016/j.cell.2007.01.015>
- Liu, J.-J., Bratkowski, M.A., Liu, X., Niu, C.-Y., Ke, A., Wang, H.-W., 2014. Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM. *Nat Struct Mol Biol* 21, 95–102. <https://doi.org/10.1038/nsmb.2736>
- Lohr, D., 1997. Nucleosome transactions on the promoters of the yeast GAL and PHO genes. *J Biol Chem* 272, 26795–26798. <https://doi.org/10.1074/jbc.272.43.26795>
- Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., Pringle, J.R., 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* 14, 953–961. [https://doi.org/10.1002/\(SICI\)1097-0061\(199807\)14:10<953::AID-YEA293>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-0061(199807)14:10<953::AID-YEA293>3.0.CO;2-U)

- Losh, J.S., King, A.K., Bakelar, J., Taylor, L., Loomis, J., Rosenzweig, J.A., Johnson, S.J., van Hoof, A., 2015. Interaction between the RNA-dependent ATPase and poly(A) polymerase subunits of the TRAMP complex is mediated by short peptides and important for snoRNA processing. *Nucleic Acids Res* 43, 1848–1858. <https://doi.org/10.1093/nar/gkv005>
- Loya, T.J., O'Rourke, T.W., Degtyareva, N., Reines, D., 2013a. A network of interdependent molecular interactions describes a higher order Nrd1-Nab3 complex involved in yeast transcription termination. *J Biol Chem* 288, 34158–34167. <https://doi.org/10.1074/jbc.M113.516765>
- Loya, T.J., O'Rourke, T.W., Reines, D., 2013b. Yeast Nab3 protein contains a self-assembly domain found in human heterogeneous nuclear ribonucleoprotein-C (hnRNP-C) that is necessary for transcription termination. *J Biol Chem* 288, 2111–2117. <https://doi.org/10.1074/jbc.M112.430678>
- Loya, T.J., O'Rourke, T.W., Reines, D., 2012. A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3. *Nucleic Acids Res* 40, 7476–7491. <https://doi.org/10.1093/nar/gks377>
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. <https://doi.org/10.1038/38444>
- Lunde, B.M., Hörner, M., Meinhart, A., 2011. Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic Acids Res.* 39, 337–346. <https://doi.org/10.1093/nar/gkq751>
- Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., Varani, G., 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 17, 1195–1201. <https://doi.org/10.1038/nsmb.1893>
- Luo, W., 2006. The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes & Development* 20, 954–965. <https://doi.org/10.1101/gad.1409106>
- Lybecker, M., Bilusic, I., Raghavan, R., 2014. Pervasive transcription: detecting functional RNAs in bacteria. *Transcription* 5, e944039. <https://doi.org/10.4161/21541272.2014.944039>
- Lyons, T.J., Gasch, A.P., Gaither, L.A., Botstein, D., Brown, P.O., Eide, D.J., 2000. Genome-wide characterization of the Zap1p zinc-responsive regulon in yeast. *Proc Natl Acad Sci U S A* 97, 7957–7962. <https://doi.org/10.1073/pnas.97.14.7957>
- MacGilvray, M.E., Shishkova, E., Place, M., Wagner, E.R., Coon, J.J., Gasch, A.P., 2020. Phosphoproteome Response to Dithiothreitol Reveals Unique Versus Shared Features of *Saccharomyces cerevisiae* Stress Responses. *J Proteome Res* 19, 3405–3417. <https://doi.org/10.1021/acs.jproteome.0c00253>
- Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., Jacquier, A., 2015. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* 4, e06722. <https://doi.org/10.7554/eLife.06722>
- Malik, S., Hisatake, K., Sumimoto, H., Horikoshi, M., Roeder, R.G., 1991. Sequence of general transcription factor TFIIB and relationships to other initiation factors. *Proceedings of the National*

Academy of Sciences 88, 9553–9557. <https://doi.org/10.1073/pnas.88.21.9553>

Malik, S., Roeder, R.G., 2000. Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends Biochem Sci* 25, 277–283. [https://doi.org/10.1016/s0968-0004\(00\)01596-6](https://doi.org/10.1016/s0968-0004(00)01596-6)

Margaritis, T., Oreal, V., Brabers, N., Maestroni, L., Vitaliano-Prunier, A., Benschop, J.J., van Hooff, S., van Leenen, D., Dargemont, C., Géli, V., Holstege, F.C.P., 2012. Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. *PLoS Genet* 8, e1002952. <https://doi.org/10.1371/journal.pgen.1002952>

Martens, J.A., 2005. Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes & Development* 19, 2695–2704. <https://doi.org/10.1101/gad.1367605>

Martens, J.A., Laprade, L., Winston, F., 2004. Intergenic transcription is required to repress the *Saccharomyces cerevisiae* SER3 gene. *Nature* 429, 571–574. <https://doi.org/10.1038/nature02538>

Martens, J.A., Wu, P.-Y.J., Winston, F., 2005. Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes Dev.* 19, 2695–2704. <https://doi.org/10.1101/gad.1367605>

Martin, B.J.E., Chruscicki, A.T., Howe, L.J., 2018. Transcription Promotes the Interaction of the FAcilitates Chromatin Transactions (FACT) Complex with Nucleosomes in *Saccharomyces cerevisiae*. *Genetics* 210, 869–881. <https://doi.org/10.1534/genetics.118.301349>

Maxon, M.E., Goodrich, J.A., Tjian, R., 1994. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes & Development* 8, 515–524. <https://doi.org/10.1101/gad.8.5.515>

Mayer, A., Heidemann, M., Lidschreiber, M., Schrieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., Cramer, P., 2012. CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. *Science* 336, 1723–1725. <https://doi.org/10.1126/science.1219651>

Meggio, F., Pinna, L.A., 2003. One-thousand-and-one substrates of protein kinase CK2? *FASEB j.* 17, 349–368. <https://doi.org/10.1096/fj.02-0473rev>

Meinhart, A., Cramer, P., 2004. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* 430, 223–226. <https://doi.org/10.1038/nature02679>

Millar, C.B., 2006. Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast. *Genes & Development* 20, 711–722. <https://doi.org/10.1101/gad.1395506>

Miller, C., Matic, I., Maier, K.C., Schwalb, B., Roether, S., Strässer, K., Tresch, A., Mann, M., Cramer, P., 2012. Mediator Phosphorylation Prevents Stress Response Transcription During Non-stress Conditions. *J. Biol. Chem.* 287, 44017–44026. <https://doi.org/10.1074/jbc.M112.430140>

Miller, M.S., Rialdi, A., Ho, J.S.Y., Tilove, M., Martinez-Gil, L., Moshkina, N.P., Peralta, Z., Noel, J., Melegari, C., Maestre, A.M., Mitsopoulos, P., Madrenas, J., Heinz, S., Benner, C., Young, J.A.T., Feagins, A.R., Basler, C.F., Fernandez-Sesma, A., Becherel, O.J., Lavin, M.F., van Bakel, H., Marazzi, I., 2015. Senataxin suppresses the antiviral transcriptional response and controls viral biogenesis. *Nat Immunol* 16, 485–494. <https://doi.org/10.1038/ni.3132>

- Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., Shilatifard, A., 2001. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A* 98, 12902–12907. <https://doi.org/10.1073/pnas.231473398>
- Mischo, H.E., Chun, Y., Harlen, K.M., Smalec, B.M., Dhir, S., Churchman, L.S., Buratowski, S., 2018. Cell-Cycle Modulation of Transcription Termination Factor Sen1. *Molecular Cell* 70, 312–326.e7. <https://doi.org/10.1016/j.molcel.2018.03.010>
- Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., Proudfoot, N.J., 2011. Yeast Sen1 Helicase Protects the Genome from Transcription-Associated Instability. *Molecular Cell* 41, 21–32. <https://doi.org/10.1016/j.molcel.2010.12.007>
- Morris, D.P., Greenleaf, A.L., 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem* 275, 39935–39943. <https://doi.org/10.1074/jbc.M004118200>
- Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., Washburn, M.P., 2009. Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. *Molecular Cell* 34, 168–178. <https://doi.org/10.1016/j.molcel.2009.02.025>
- Myers, L.C., Gustafsson, C.M., Bushnell, D.A., Lui, M., Erdjument-Bromage, H., Tempst, P., Kornberg, R.D., 1998. The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev* 12, 45–54. <https://doi.org/10.1101/gad.12.1.45>
- Nadal-Ribelles, M., Solé, C., Xu, Z., Steinmetz, L.M., de Nadal, E., Posas, F., 2014a. Control of Cdc28 CDK1 by a stress-induced lncRNA. *Mol Cell* 53, 549–561. <https://doi.org/10.1016/j.molcel.2014.01.006>
- Nadal-Ribelles, M., Solé, C., Xu, Z., Steinmetz, L.M., de Nadal, E., Posas, F., 2014b. Control of Cdc28 CDK1 by a Stress-Induced lncRNA. *Molecular Cell* 53, 549–561. <https://doi.org/10.1016/j.molcel.2014.01.006>
- Naville, M., Gautheret, D., 2009. Transcription attenuation in bacteria: theme and variations. *Briefings in Functional Genomics and Proteomics* 8, 482–492. <https://doi.org/10.1093/bfpg/elp025>
- Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., Nagy, P.L., 2008. The Glc7 Phosphatase Subunit of the Cleavage and Polyadenylation Factor Is Essential for Transcription Termination on snoRNA Genes. *Molecular Cell* 29, 577–587. <https://doi.org/10.1016/j.molcel.2007.12.031>
- Neil, Helen, Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., Jacquier, A., 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038–1042. <https://doi.org/10.1038/nature07747>
- Nemec, C.M., Singh, A.K., Ali, A., Tseng, S.C., Syal, K., Ringelberg, K.J., Ho, Y.-H., Hintermair, C., Ahmad, M.F., Kar, R.K., Gasch, A.P., Akhtar, M.S., Eick, D., Ansari, A.Z., 2019. Noncanonical CTD kinases regulate RNA polymerase II in a gene-class-specific manner. *Nat Chem Biol* 15, 123–131. <https://doi.org/10.1038/s41589-018-0194-1>
- Nevers, A., Doyen, A., Malabat, C., Néron, B., Kergrohen, T., Jacquier, A., Badis, G., 2018. Antisense transcriptional interference mediates condition-specific gene repression in budding yeast.

Nucleic Acids Res 46, 6009–6025. <https://doi.org/10.1093/nar/gky342>

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., Weissman, J.S., 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840–846. <https://doi.org/10.1038/nature04785>

Ng, H.H., Robert, F., Young, R.A., Struhl, K., 2003. Targeted Recruitment of Set1 Histone Methylase by Elongating Pol II Provides a Localized Mark and Memory of Recent Transcriptional Activity. *Molecular Cell* 11, 709–719. [https://doi.org/10.1016/S1097-2765\(03\)00092-3](https://doi.org/10.1016/S1097-2765(03)00092-3)

Nigavekar, S.S., Tan, Y.S.H., Cannon, J.F., 2002. Glc8 is a glucose-repressible activator of Glc7 protein phosphatase-1. *Arch Biochem Biophys* 404, 71–79. [https://doi.org/10.1016/s0003-9861\(02\)00231-x](https://doi.org/10.1016/s0003-9861(02)00231-x)

Ofiteru, A.M., Ruta, L.L., Rotaru, C., Dumitru, I., Ene, C.D., Neagoe, A., Farcasanu, I.C., 2012. Overexpression of the PHO84 gene causes heavy metal accumulation and induces Ire1p-dependent unfolded protein response in *Saccharomyces cerevisiae* cells. *Appl Microbiol Biotechnol* 94, 425–435. <https://doi.org/10.1007/s00253-011-3784-3>

Ogami, K., Chen, Y., Manley, J.L., 2018. RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Noncoding RNA* 4. <https://doi.org/10.3390/ncrna4010008>

Pagani, M.A., Casamayor, A., Serrano, R., Atrian, S., Ariño, J., 2007. Disruption of iron homeostasis in *Saccharomyces cerevisiae* by high zinc levels: a genome-wide study. *Mol Microbiol* 65, 521–537. <https://doi.org/10.1111/j.1365-2958.2007.05807.x>

Pal, M., Luse, D.S., 2003. The initiation-elongation transition: Lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23. *Proceedings of the National Academy of Sciences* 100, 5700–5705. <https://doi.org/10.1073/pnas.1037057100>

Park, J., Kang, M., Kim, M., 2015. Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1. *Nucleic Acids Res* 43, 2625–2637. <https://doi.org/10.1093/nar/gkv133>

Pearson, E., Moore, C., 2014. The Evolutionarily Conserved Pol II Flap Loop Contributes to Proper Transcription Termination on Short Yeast Genes. *Cell Reports* 9, 821–828. <https://doi.org/10.1016/j.celrep.2014.10.007>

Pearson, E.L., Moore, C.L., 2013. Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor Rat1. *J Biol Chem* 288, 19750–19759. <https://doi.org/10.1074/jbc.M112.434985>

Pelechano, V., Steinmetz, L.M., 2013. Gene regulation by antisense transcription. *Nat. Rev. Genet.* 14, 880–893. <https://doi.org/10.1038/nrg3594>

Porrua, O., Boudvillain, M., Libri, D., 2016. Transcription Termination: Variations on Common Themes. *Trends Genet* 32, 508–522. <https://doi.org/10.1016/j.tig.2016.05.007>

Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R., Libri, D., 2012. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* 31, 3935–3948. <https://doi.org/10.1038/emboj.2012.237>



- Porrúa, O., Libri, D., 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* 16, 190–202. <https://doi.org/10.1038/nrm3943>
- Porrúa, O., Libri, D., 2015. Characterization of the mechanisms of transcription termination by the helicase Sen1. *Methods Mol. Biol.* 1259, 313–331. [https://doi.org/10.1007/978-1-4939-2214-7\\_19](https://doi.org/10.1007/978-1-4939-2214-7_19)
- Porrúa, O., Libri, D., 2013. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat Struct Mol Biol* 20, 884–891. <https://doi.org/10.1038/nsmb.2592>
- Poss, Z.C., Ebmeier, C.C., Taatjes, D.J., 2013. The Mediator complex and transcription regulation. *Critical Reviews in Biochemistry and Molecular Biology* 48, 575–608. <https://doi.org/10.3109/10409238.2013.840259>
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., Jensen, T.H., 2008. RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science* 322, 1851–1854. <https://doi.org/10.1126/science.1164096>
- Qiu, H., Hu, C., 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. <https://doi.org/10.1016/j.molcel.2009.02.018>
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., Manke, T., 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160-165. <https://doi.org/10.1093/nar/gkw257>
- Rando, O.J., Winston, F., 2012. Chromatin and Transcription in Yeast. *Genetics* 190, 351–387. <https://doi.org/10.1534/genetics.111.132266>
- Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W., Lewis, B.A., 2012. Evidence of the Involvement of O -GlcNAc-modified Human RNA Polymerase II CTD in Transcription *in Vitro* and *in Vivo*. *J. Biol. Chem.* 287, 23549–23561. <https://doi.org/10.1074/jbc.M111.330910>
- Rawal, C.C., Zardoni, L., Di Terlizzi, M., Galati, E., Brambati, A., Lazzaro, F., Liberi, G., Pelliccioli, A., 2020. Senataxin Ortholog Sen1 Limits DNA:RNA Hybrid Accumulation at DNA Double-Strand Breaks to Control End Resection and Repair Fidelity. *Cell Reports* 31, 107603. <https://doi.org/10.1016/j.celrep.2020.107603>
- Rhee, H.S., Pugh, B.F., 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 483, 295–301. <https://doi.org/10.1038/nature10799>
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., Séraphin, B., 1999. A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* 17, 1030–1032. <https://doi.org/10.1038/13732>
- Robinson, P.J., Trnka, M.J., Bushnell, D.A., Davis, R.E., Mattei, P.-J., Burlingame, A.L., Kornberg, R.D., 2016. Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* 166, 1411-1422.e16. <https://doi.org/10.1016/j.cell.2016.08.050>
- Rodríguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L., Buratowski, S., 2000. Kin28, the TFIIF-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA

processing machinery to RNA polymerase II. *Mol Cell Biol* 20, 104–112.  
<https://doi.org/10.1128/mcb.20.1.104-112.2000>

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W.W., Wilm, M., Aasland, R., Stewart, A.F., 2001. The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J* 20, 7137–7148. <https://doi.org/10.1093/emboj/20.24.7137>

Rolland, F., Winderickx, J., Thevelein, J.M., 2002. Glucose-sensing and -signalling mechanisms in yeast. *FEMS Yeast Research* 2, 183–201. <https://doi.org/10.1111/j.1567-1364.2002.tb00084.x>

Sainsbury, S., Bernecky, C., Cramer, P., 2015. Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Biol* 16, 129–143. <https://doi.org/10.1038/nrm3952>

Saunders, A., Core, L.J., Lis, J.T., 2006. Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* 7, 557–567. <https://doi.org/10.1038/nrm1981>

Sawadogo, M., Roeder, R.G., 1985. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proceedings of the National Academy of Sciences* 82, 4394–4398. <https://doi.org/10.1073/pnas.82.13.4394>

Sayou, C., Millán-Zambrano, G., Santos-Rosa, H., Petfalski, E., Robson, S., Houseley, J., Kouzarides, T., Tollervey, D., 2017. RNA Binding by Histone Methyltransferases Set1 and Set2. *Mol Cell Biol* 37, e00165-17, e00165-17. <https://doi.org/10.1128/MCB.00165-17>

Schaughency, P., Merran, J., Corden, J.L., 2014. Genome-Wide Mapping of Yeast RNA Polymerase II Termination. *PLoS Genet* 10, e1004632.  
<https://doi.org/10.1371/journal.pgen.1004632>

Schmidt, K., Butler, J.S., 2013. Nuclear RNA surveillance: role of TRAMP in controlling exosome specificity. *Wiley Interdiscip Rev RNA* 4, 217–231. <https://doi.org/10.1002/wrna.1155>

Schneider, C., Tollervey, D., 2013. Threading the barrel of the RNA exosome. *Trends Biochem Sci* 38, 485–493. <https://doi.org/10.1016/j.tibs.2013.06.013>

Schreieck, A., Easter, A.D., Etzold, S., Wiederhold, K., Lidschreiber, M., Cramer, P., 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.  
<https://doi.org/10.1038/nsmb.2753>

Schröder, S., Herker, E., Itzen, F., He, D., Thomas, S., Gilchrist, D.A., Kaehlcke, K., Cho, S., Pollard, K.S., Capra, J.A., Schnölzer, M., Cole, P.A., Geyer, M., Bruneau, B.G., Adelman, K., Ott, M., 2013. Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. *Mol Cell* 52, 314–324. <https://doi.org/10.1016/j.molcel.2013.10.009>

Schroeder, S.C., Schwer, B., Shuman, S., Bentley, D., 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* 14, 2435–2440.  
<https://doi.org/10.1101/gad.836300>

Schuch, B., Feigenbutz, M., Makino, D.L., Falk, S., Basquin, C., Mitchell, P., Conti, E., 2014. The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J* 33, 2829–2846. <https://doi.org/10.15252/embj.201488757>

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., Cramer, P., 2013a. Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell*

155, 1075–1087. <https://doi.org/10.1016/j.cell.2013.10.024>

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., Cramer, P., 2013b. Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell*. <https://doi.org/10.1016/j.cell.2013.10.024>

Schweikhard, V., Meng, C., Murakami, K., Kaplan, C.D., Kornberg, R.D., Block, S.M., 2014. Transcription factors TFIIF and TFIIS promote transcript elongation by RNA polymerase II by synergistic and independent mechanisms. *Proceedings of the National Academy of Sciences* 111, 6642–6647. <https://doi.org/10.1073/pnas.1405181111>

Selth, L.A., Sigurdsson, S., Svejstrup, J.Q., 2010. Transcript Elongation by RNA Polymerase II. *Annu. Rev. Biochem.* 79, 271–293. <https://doi.org/10.1146/annurev.biochem.78.062807.091425>

Shen, A., Lupardus, P.J., Morell, M., Ponder, E.L., Sadaghiani, A.M., Garcia, K.C., Bogoy, M., 2009. Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. *PLoS One* 4, e8119. <https://doi.org/10.1371/journal.pone.0008119>

Sigurdsson, S., Dirac-Svejstrup, A.B., Svejstrup, J.Q., 2010. Evidence that Transcript Cleavage Is Essential for RNA Polymerase II Transcription and Cell Viability. *Molecular Cell* 38, 202–210. <https://doi.org/10.1016/j.molcel.2010.02.026>

Sims, R.J., 2004. Elongation by RNA polymerase II: the short and long of it. *Genes & Development* 18, 2437–2468. <https://doi.org/10.1101/gad.1235904>

Sims, R.J., Rojas, L.A., Beck, D., Bonasio, R., Schuller, R., Drury, W.J., Eick, D., Reinberg, D., 2011. The C-Terminal Domain of RNA Polymerase II Is Modified by Site-Specific Methylation. *Science* 332, 99–103. <https://doi.org/10.1126/science.1202663>

Singleton, M.R., Dillingham, M.S., Wigley, D.B., 2007. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76, 23–50. <https://doi.org/10.1146/annurev.biochem.76.052305.115300>

Skourti-Stathaki, K., Proudfoot, N.J., Gromak, N., 2011. Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Molecular Cell* 42, 794–805. <https://doi.org/10.1016/j.molcel.2011.04.026>

Smale, S.T., Kadonaga, J.T., 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* 72, 449–479. <https://doi.org/10.1146/annurev.biochem.72.121801.161520>

Soudet, J., Gill, J.K., Stutz, F., 2018. Noncoding transcription influences the replication initiation program through chromatin regulation. *Genome Res* 28, 1882–1893. <https://doi.org/10.1101/gr.239582.118>

Soutourina, J., 2018. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol* 19, 262–274. <https://doi.org/10.1038/nrm.2017.115>

Soutourina, J., Wydau, S., Ambroise, Y., Boschiero, C., Werner, M., 2011. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* 331, 1451–1454. <https://doi.org/10.1126/science.1200188>

Spain, M.M., Ansari, S.A., Pathak, R., Palumbo, M.J., Morse, R.H., Govind, C.K., 2014. The RSC Complex Localizes to Coding Sequences to Regulate Pol II and Histone Occupancy. *Molecular Cell*

56, 653–666. <https://doi.org/10.1016/j.molcel.2014.10.002>

Squazzo, S.L., Costa, P.J., Lindstrom, D.L., Kumer, K.E., Simic, R., Jennings, J.L., Link, A.J., Arndt, K.M., Hartzog, G.A., 2002. The Paf1 complex physically and functionally associates with transcription elongation factors in vivo. *EMBO J* 21, 1764–1774. <https://doi.org/10.1093/emboj/21.7.1764>

Steinmetz, E.J., Conrad, N.K., Brow, D.A., Corden, J.L., 2001. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413, 327–331. <https://doi.org/10.1038/35095090>

Struhl, K., 1999. Fundamentally Different Logic of Gene Regulation in Eukaryotes and Prokaryotes. *Cell* 98, 1–4. [https://doi.org/10.1016/S0092-8674\(00\)80599-1](https://doi.org/10.1016/S0092-8674(00)80599-1)

Studier, F.W., 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234.

Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J., Lavin, M.F., 2009. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Human Molecular Genetics* 18, 3384–3396. <https://doi.org/10.1093/hmg/ddp278>

Svejstrup, J.Q., 2007. Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem Sci* 32, 165–171. <https://doi.org/10.1016/j.tibs.2007.02.005>

Swaney, Danielle L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J., Villén, J., 2013. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat. Methods* 10, 676–682. <https://doi.org/10.1038/nmeth.2519>

Taatjes, D.J., 2017. The Continuing SAGA of TFIID and RNA Polymerase II Transcription. *Molecular Cell* 68, 1–2. <https://doi.org/10.1016/j.molcel.2017.09.028>

Tan, S., Aso, T., Conaway, R.C., Conaway, J.W., 1994. Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. *J. Biol. Chem.* 269, 25684–25691.

Tang, L., Nogales, E., Ciferri, C., 2010. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Prog Biophys Mol Biol* 102, 122–128. <https://doi.org/10.1016/j.pbiomolbio.2010.05.001>

Theisen, J.W.M., Lim, C.Y., Kadonaga, J.T., 2010. Three Key Subregions Contribute to the Function of the Downstream RNA Polymerase II Core Promoter. *MCB* 30, 3471–3479. <https://doi.org/10.1128/MCB.00053-10>

Thiebaut, M., Colin, J., Neil, H., Jacquier, A., Séraphin, B., Lacroute, F., Libri, D., 2008. Futile cycle of transcription initiation and termination modulates the response to nucleotide shortage in *S. cerevisiae*. *Mol. Cell* 31, 671–682. <https://doi.org/10.1016/j.molcel.2008.08.010>

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., Libri, D., 2006a. Transcription Termination and Nuclear Degradation of Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. *Molecular Cell* 23, 853–864. <https://doi.org/10.1016/j.molcel.2006.07.029>

Thompson, C.M., Young, R.A., 1995. General requirement for RNA polymerase II holoenzymes in vivo. *Proc Natl Acad Sci U S A* 92, 4587–4590. <https://doi.org/10.1073/pnas.92.10.4587>

- Toesca, I., Nery, C.R., Fernandez, C.F., Sayani, S., Chanfreau, G.F., 2011. Cryptic transcription mediates repression of subtelomeric metal homeostasis genes. *PLoS Genet* 7, e1002163. <https://doi.org/10.1371/journal.pgen.1002163>
- Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., Libri, D., 2014. Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Molecular Cell* 55, 467–481. <https://doi.org/10.1016/j.molcel.2014.05.031>
- Tudek, A., Schmid, M., Makaras, M., Barrass, J.D., Beggs, J.D., Jensen, T.H., 2018. A Nuclear Export Block Triggers the Decay of Newly Synthesized Polyadenylated RNA. *Cell Rep* 24, 2457–2467.e7. <https://doi.org/10.1016/j.celrep.2018.07.103>
- Ursic, D., Chinchilla, K., Finkel, J.S., Culbertson, M.R., 2004. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res* 32, 2441–2452. <https://doi.org/10.1093/nar/gkh561>
- van Dijk, E.L., Chen, C.L., d'Aubenton-Carafa, Y., Gourvenec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., Loeillet, S., Nicolas, A., Thermes, C., Morillon, A., 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475, 114–117. <https://doi.org/10.1038/nature10118>
- Van Oss, S.B., Carvunis, A.-R., 2019. De novo gene birth. *PLoS Genet* 15, e1008160. <https://doi.org/10.1371/journal.pgen.1008160>
- van Werven, F.J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., Amon, A., 2012. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* 150, 1170–1181. <https://doi.org/10.1016/j.cell.2012.06.049>
- Vanacova, S., Stefl, R., 2007. The exosome and RNA quality control in the nucleus. *EMBO Rep* 8, 651–657. <https://doi.org/10.1038/sj.embor.7401005>
- Vanáčová, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., Keller, W., 2005. A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol* 3, e189. <https://doi.org/10.1371/journal.pbio.0030189>
- van Werven, F.J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., Amon, A., 2012. Transcription of Two Long Noncoding RNAs Mediates Mating-Type Control of Gametogenesis in Budding Yeast. *Cell* 150, 1170–1181. <https://doi.org/10.1016/j.cell.2012.06.049>
- Varet, H., Brillet-Guéguen, L., Coppée, J.-Y., Dillies, M.-A., 2016. SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS ONE* 11, e0157022. <https://doi.org/10.1371/journal.pone.0157022>
- Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., Meinhart, A., 2008. The Nrd1–Nab3–Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* 15, 795–804. <https://doi.org/10.1038/nsmb.1468>
- Verrijzer, C.P., Chen, J.L., Yokomori, K., Tjian, R., 1995. Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II. *Cell* 81, 1115–1125. [https://doi.org/10.1016/s0092-8674\(05\)80016-9](https://doi.org/10.1016/s0092-8674(05)80016-9)

- Villa, T., Barucco, M., Martin-Niclos, M.-J., Jacquier, A., Libri, D., 2020. Degradation of Non-coding RNAs Promotes Recycling of Termination Factors at Sites of Transcription. *Cell Rep* 32, 107942. <https://doi.org/10.1016/j.celrep.2020.107942>
- Wade, J.T., Grainger, D.C., 2014. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol* 12, 647–653. <https://doi.org/10.1038/nrmicro3316>
- Wakula, P., Beullens, M., Ceulemans, H., Stalmans, W., Bollen, M., 2003. Degeneracy and Function of the Ubiquitous RV X F Motif That Mediates Binding to Protein Phosphatase-1. *J. Biol. Chem.* 278, 18817–18823. <https://doi.org/10.1074/jbc.M300175200>
- Wang, S., Han, Z., Libri, D., Porrua, O., Strick, T.R., 2019. Single-molecule characterization of extrinsic transcription termination by Sen1 helicase. *Nat Commun* 10, 1545. <https://doi.org/10.1038/s41467-019-09560-9>
- Wang, Y., Weisenhorn, E., MacDiarmid, C.W., Andreini, C., Bucci, M., Taggart, J., Banci, L., Russell, J., Coon, J.J., Eide, D.J., 2018. The cellular economy of the *Saccharomyces cerevisiae* zinc proteome. *Metallomics* 10, 1755–1776. <https://doi.org/10.1039/c8mt00269j>
- Warfield, L., Ramachandran, S., Baptista, T., Devys, D., Tora, L., Hahn, S., 2017. Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID. *Molecular Cell* 68, 118-129.e5. <https://doi.org/10.1016/j.molcel.2017.08.014>
- Wasmuth, E.V., Januszyk, K., Lima, C.D., 2014. Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. *Nature* 511, 435–439. <https://doi.org/10.1038/nature13406>
- Wasmuth, E.V., Zinder, J.C., Zattas, D., Das, M., Lima, C.D., 2017. Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that Mpp6 stimulates RNA decay and recruits the Mtr4 helicase. *Elife* 6. <https://doi.org/10.7554/eLife.29062>
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Weake, V.M., Workman, J.L., 2012. SAGA function in tissue-specific gene expression. *Trends Cell Biol* 22, 177–184. <https://doi.org/10.1016/j.tcb.2011.11.005>
- Webb, S., Hector, R.D., Kudla, G., Granneman, S., 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* 15, R8. <https://doi.org/10.1186/gb-2014-15-1-r8>
- Weir, J.R., Bonneau, F., Hentschel, J., Conti, E., 2010. Structural analysis reveals the characteristic features of Mtr4, a DEXH helicase involved in nuclear RNA processing and surveillance. *Proc Natl Acad Sci U S A* 107, 12139–12144. <https://doi.org/10.1073/pnas.1004953107>
- Wilson, M.D., Harreman, M., Svejstrup, J.Q., 2013. Ubiquitylation and degradation of elongating RNA polymerase II: the last resort. *Biochim Biophys Acta* 1829, 151–157. <https://doi.org/10.1016/j.bbagr.2012.08.002>
- Wilson, S., Bird, A.J., 2016. Zinc sensing and regulation in yeast model systems. *Arch Biochem Biophys* 611, 30–36. <https://doi.org/10.1016/j.abb.2016.02.031>

- Wilson, S.M., Datar, K.V., Paddy, M.R., Swedlow, J.R., Swanson, M.S., 1994. Characterization of nuclear polyadenylated RNA-binding proteins in *Saccharomyces cerevisiae*. *The Journal of Cell Biology* 127, 1173–1184. <https://doi.org/10.1083/jcb.127.5.1173>
- Wlotzka, W., Kudla, G., Granneman, S., Tollervey, D., 2011. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts: Targets for nuclear surveillance. *The EMBO Journal* 30, 1790–1803. <https://doi.org/10.1038/emboj.2011.97>
- Wong, C.-M., Kewu, P., Zhe, H., Lee, J.T., 2015. Current perspectives on the role of TRAMP in nuclear RNA surveillance and quality control. *RRBC* 111. <https://doi.org/10.2147/RRBC.S58509>
- Wong, K.H., Jin, Y., Struhl, K., 2014. TFIIF phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Mol Cell* 54, 601–612. <https://doi.org/10.1016/j.molcel.2014.03.024>
- Woo, H., Dam Ha, S., Lee, S.B., Buratowski, S., Kim, T., 2017. Modulation of gene expression dynamics by co-transcriptional histone methylations. *Exp Mol Med* 49, e326–e326. <https://doi.org/10.1038/emm.2017.19>
- Wu, C.-Y., Bird, A.J., Chung, L.M., Newton, M.A., Winge, D.R., Eide, D.J., 2008. Differential control of Zap1-regulated genes in response to zinc deficiency in *Saccharomyces cerevisiae*. *BMC Genomics* 9, 370. <https://doi.org/10.1186/1471-2164-9-370>
- Wyce, A., Xiao, T., Whelan, K.A., Kosman, C., Walter, W., Eick, D., Hughes, T.R., Krogan, N.J., Strahl, B.D., Berger, S.L., 2007. H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. *Mol Cell* 27, 275–288. <https://doi.org/10.1016/j.molcel.2007.01.035>
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., Libri, D., Jacquier, A., 2005. Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. *Cell* 121, 725–737. <https://doi.org/10.1016/j.cell.2005.04.030>
- Xiao, T., 2003. Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. *Genes & Development* 17, 654–663. <https://doi.org/10.1101/gad.1055503>
- Xu, Zhenyu, Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., Steinmetz, L.M., 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033–1037. <https://doi.org/10.1038/nature07728>
- Yaakov, G., Duch, A., García-Rubio, M., Clotet, J., Jimenez, J., Aguilera, A., Posas, F., 2009. The Stress-activated Protein Kinase Hog1 Mediates S Phase Delay in Response to Osmostress. *MBoC* 20, 3572–3582. <https://doi.org/10.1091/mbc.e09-02-0129>
- Yang, C., Bolotin, E., Jiang, T., Sladek, F.M., Martinez, E., 2007. 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* 389, 52–65. <https://doi.org/10.1016/j.gene.2006.09.029>
- Yoon, S., Qiu, H., Swanson, M.J., Hinnebusch, A.G., 2003. Recruitment of SWI/SNF by Gcn4p does not require Snf2p or Gcn5p but depends strongly on SWI/SNF integrity, SRB mediator, and SAGA. *Mol Cell Biol* 23, 8829–8845. <https://doi.org/10.1128/mcb.23.23.8829-9945.2003>
- Yu, Y., Yarrington, R.M., Chuong, E.B., Elde, N.C., Stillman, D.J., 2016. Disruption of promoter

memory by synthesis of a long noncoding RNA. *Proc Natl Acad Sci USA* 113, 9575–9580. <https://doi.org/10.1073/pnas.1601793113>

Yurko, N., Liu, X., Yamazaki, T., Hoque, M., Tian, B., Manley, J.L., 2017. MPK1/SLT2 Links Multiple Stress Responses with Gene Expression in Budding Yeast by Phosphorylating Tyr1 of the RNAP II CTD. *Molecular Cell* 68, 913-925.e3. <https://doi.org/10.1016/j.molcel.2017.11.020>

Zhang, Y., Chun, Y., Buratowski, S., Tong, L., 2019. Identification of Three Sequence Motifs in the Transcription Termination Factor Sen1 that Mediate Direct Interactions with Nrd1. *Structure* 27, 1156-1161.e4. <https://doi.org/10.1016/j.str.2019.04.005>

Zhao, H., Butler, E., Rodgers, J., Spizzo, T., Duesterhoeft, S., Eide, D., 1998. Regulation of zinc homeostasis in yeast by binding of the ZAP1 transcriptional activator to zinc-responsive promoter elements. *J Biol Chem* 273, 28713–28720. <https://doi.org/10.1074/jbc.273.44.28713>

Zhao, Y.-Y., Cao, C.-L., Liu, Y.-L., Wang, J., Li, J., Li, S.-Y., Deng, Y., 2020. Identification of the Genetic Requirements for Zinc Tolerance and Toxicity in *Saccharomyces cerevisiae*. *G3 (Bethesda)* 10, 479–488. <https://doi.org/10.1534/g3.119.400933>

Zinder, J.C., Lima, C.D., 2017. Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. *Genes Dev* 31, 88–100. <https://doi.org/10.1101/gad.294769.116>

Zinder, J.C., Wasmuth, E.V., Lima, C.D., 2016. Nuclear RNA Exosome at 3.1 Å Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3. *Mol Cell* 64, 734–745. <https://doi.org/10.1016/j.molcel.2016.09.038>



