Rôle de la transcription pervasive antisens chez Saccharomyces cerevisiae dans la régulation de l'expression des gènes

Alicia Chery

▶ **To cite this version:**

Alicia Chery. Rôle de la transcription pervasive antisens chez Saccharomyces cerevisiae dans la régulation de l'expression des gènes. Biologie moléculaire. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066191 . tel-01662409

HAL Id: tel-01662409
https://theses.hal.science/tel-01662409

Submitted on 13 Dec 2017

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.

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant (ED 515)

Laboratoire de Génétique des Interactions Macromoléculaires CNRS UMR 3525

Rôle de la transcription pervasive antisens chez *Saccharomyces cerevisiae* dans la régulation de l'expression des gènes

Par Alicia CHERY (Epouse NEVERS)

Thèse de doctorat de Biologie Moléculaire et Cellulaire

Dirigée par Alain JACQUIER et Gwenaël BADIS-BREARD

Présentée et soutenue publiquement le 04 Octobre 2017

Devant un jury composé de :

Mr Domenico LIBRI – Président du Jury / Rapporteur

Mme Françoise STUTZ – Rapporteur

Mr Damien BREGEON – Examinateur

Mme Julie SOUTOURINA – Examinateur

Mme Gwenaël BADIS-BREARD – Co-encadrant de thèse

Mr Alain JACQUIER – Directeur de thèse

À Quentin Nevers, ma plus grande force,

À Hugo Nevers, la plus belle des motivations…

Remerciements

Merci à Françoise Stutz, Domenico Libri, Damien Bregeon et Julie Soutourina d'avoir accepté de faire partie de mon jury de thèse, c'est un honneur pour moi d'échanger avec vous et de vous soumettre le travail que j'ai réalisé pendant ma thèse.

Un grand merci à Alain Jacquier, mon directeur de thèse, de m'avoir accueillie, accompagnée et formée au sein de son laboratoire de recherche. Les interactions que j'ai pu avoir avec toi m'ont grandement enrichie sur le plan scientifique, et m'ont souvent poussée à dépasser mes appréhensions et mon manque d'assurance pour mener à bien les projets dans lesquels tu m'as impliquée.

Merci à Gwenaël Badis-Bréard, mon encadrante de thèse, d'avoir accepté de me prendre sous sa responsabilité après ma première année de thèse. Tu as su me pousser et me motiver dans les moments difficiles, et tu m'as accordé ta confiance en me laissant prendre mon indépendance quand j'en ressentais le besoin. Je te suis reconnaissante pour l'accompagnement scientifique que tu m'as offert jusqu'ici et pour l'attention que tu as porté à mon bien-être personnel pendant ma thèse.

Merci à Frank Feuerbach pour son encadrement lors de mon stage de Master 2 et pendant ma première année de thèse. Cela a été un plaisir de faire ta connaissance, et une chance de profiter de ta formation aux expérimentations en laboratoire, tant tu m'impressionnes par la quantité de choses que tu maîtrises.

Antonia je te remercie pour toute la quantité de travail que tu as fourni pour m'aider dans ma thèse, j'ai adoré travailler avec toi, et pour moi tu resteras toujours la déesse de l'extraction des ARN ! Tu as toujours été là pour moi, dans les moments heureux ou plus durs, et tu es la voisine de paillasse rêvée !

Merci à Christophe Malabat et Thomas Kergrohen pour leur travail dans le projet de ma thèse. Merci aussi pour leur aide, leur disponibilité et leur bienveillance quand je me suis initiée à l'analyse des données.

Elodie, mon amie et collègue, ça a été un privilège de vivre cette thèse à tes côtés. On s'est mutuellement soutenues, et notre amitié en est d'autant plus forte aujourd'hui ! Je te souhaite le meilleur, tant au niveau professionnel que personnel. Ça semble bien parti pour, et de toute façon je serai là pour voir ça !

Ma petite Lucia, toi qui m'a vue arriver toute timide et perdue en janvier 2013, tu es devenue si importante aujourd'hui. Je n'imagine pas ce que ça aurait été au labo sans toi qui a toujours été là pour me remonter le moral, et me changer les idées. Merci…

Je remercie les membres actuels de mon laboratoire pour leur présence, leur gentillesse, leur soutien et leur aide tout au long de mon parcours. Abdelkader Namane, Cosmin Saveanu, Laurence Decourty, Marine Dehecq, Christelle Lenormand et Claudine Porée, je garderai toujours un bon souvenir du laboratoire grâce à vous. Je souhaite remercier particulièrement Micheline Fromont-Racine qui a toujours été présente chaque fois que j'en ai eu besoin notamment pour m'avancer des frais de congrès, ou même pour une lettre de recommandation lors de mes recherches d'appartement. Merci aussi aux membres passés, au risque d'en oublier, comme John Mouaikel, Quentin Defenouillère, Salomé Joinnet, Monique Moulin et Birame Name…. Merci aussi aux membres de la plate-forme de transcriptomique, notamment Caroline Proux et Odile Sismeiro pour leur accessibilité, et leur gentillesse.

Merci à tous les collègues du 4^{e} étage du bâtiment Fernbach, dont la présence a parfois été vitale pour supporter certaines journées : Simon Malesys, Lucie Poggi, Rostyslav Makarenko, Célia Raimondi, Frédéric Moyrand et Samia Miled je vous souhaite plein de réussite !

Valentine Mosbach, on en a fait du chemin depuis le M2 ! Je te remercie pour les bons moments passés en ta compagnie, pour le super voyage en Savoie, et pour toutes les petites attentions que tu eu pour moi.

De façon plus générale, je remercie tous les membres des équipes de Christophe D'Enfert, Guilem Janbon, Guy-Frank Richard et Benoît Arcangioli, qui m'ont, à un moment ou un autre, dépannée d'un réactif ou aidé pour une question technique.

Merci à tous mes amis, qui m'ont permis de garder une vie sociale « normale » et de passer de très bons moments pendant ces quatre dernières années, en particulier Tony, Lucie, Antonin, Gautier, Carine et Monique. Je suis très heureuse de partager l'aboutissement de ce travail avec vous.

J'en arrive à ma famille, qui m'a soutenue depuis le début, et qui a tout fait pour que je me sente bien dans ma thèse, en m'aidant à m'épanouir à l'extérieur. Mille mercis à mes frère et sœurs Lauren, Melvyn et Naomy qui par leur amour et leurs encouragements m'ont permis d'en arriver là aujourd'hui. Merci infiniment à Noëlle-Anne, Jean-Pierre et Julie, de m'avoir aussi spontanément accueillie et considérée comme un membre de la famille. Je vous remercie d'avoir été présents et de m'avoir soutenue dans les projets professionnels ou personnels avec toute la bienveillance du monde.

Pour finir, je souhaite remercier mon mari, Quentin, qui a été incroyable avec moi depuis qu'on se connaît. Tu m'as aidé à croire en moi et en mes capacités, à prendre de l'assurance et tu m'as toujours encouragée quand j'en avais le plus besoin. J'ai vraiment eu de la chance de faire ma thèse en même temps que toi, de profiter de tes conseils, de nos discussions scientifiques et surtout de l'immense bonheur qui tu m'apportes au quotidien.

Résumé

L'expression des gènes est finement régulée dans la cellule et soumise à de multiples contrôles-qualité. Cette régulation intervient à différents niveaux, de façon à garantir une synthèse efficace des produits fonctionnels de l'expression génique, et pour assurer une adaptation à un changement environnemental. Notamment, les régulations transcriptionnelles sont cruciales pour contrôler la cinétique et le niveau d'expression des gènes. La transcription pervasive est une transcription généralisée non-codante et instable qui fut révélée chez la levure *Saccharomyces cerevisiae*. Bien que son potentiel régulateur ait été démontré de façon ponctuelle, la question de sa fonctionnalité globale restait ouverte. Lors de ma thèse, j'ai pu montrer l'existence de phénomènes multiples d'interférence transcriptionnelle liés à la transcription pervasive, pour co-réguler un ensemble de gènes entre la phase exponentielle et la quiescence. En effet, la transcription non-codante en antisens des gènes concernés conduit à leur répression, dans des conditions où ils ne doivent pas être exprimés. Le mécanisme de répression fait intervenir des modifications de la chromatine. La levure bourgeonnante, dépourvue de la machinerie d'ARN interférence, présente donc un système fin de régulation de l'expression génique utilisant la transcription pervasive.

Abstract

In the cell, gene expression is finely tuned and is submitted to different quality-controls. Gene are regulated at different expression levels in order to guarantee a proper synthesis of functional products, and to ensure an optimal adaptation to environmental changes. In particular, transcriptional regulations are critical for gene expression level and kinetics.

Pervasive transcription, defined as a generalized non-coding and unstable transcription, was discovered in the yeast *Saccharomyces cerevisiae.* Although its regulatory potential was punctually shown, the question of its global functionality still remained. During my PhD, I could show the existence of numerous transcriptional interference mechanisms involved in the co-regulation of a group of genes between exponential phase and quiescence. Indeed, non-coding transcription in antisense to genes promoter leads to its repression in conditions where they have to be switched off. The repression mechanism is allowed by chromatin modifications.

Hence, budding yeast that lacks RNA interference machinery has developed a fine regulation system using pervasive transcription.

Liste des abréviations

- ADN : Acide Désoxyribonucléique
- ARN : Acide Ribonucléique
- ARNm : ARN messager
- ARN pol : ARN polymérase
- siARN: petit ARN interférents, « small interfering RNA »
- miARN: micro ARN, « microRNA »
- Mb: mégabases
- PIC : Complexe de Pré-initiation, « Pre-Initiation Complex"
- GTF : Facteurs généraux de transcription, « General Transcription Factors »
- CTD : Domaine C-terminal, « C-terminal domain »
- TBP : Protéine liée à l'élément TATA, « TATA-binding protein »
- ORF : Phase ouverte de lecture, « Open-Reading frame »
- UTR : Région non transcrite, « Untranslated region »
- NFR : Région pauvre en nucléosomes, « Nucleosome-free region »
- EF : Facteurs d'élongation transcriptionnelle, « Elongation Factors »
- HDAC : Histone-déacétylase
- HMT : Histone-méthyl-transférase
- ncARN : ARN non-codant
- lncARN : longs ARN non-codants
- NMD : Voie de dégradation des ARNm non-sens, « Nonsense-mediated mRNA decay »
- CUT : Transcrits cryptiques instables, « Cryptic Unstable Transcripts »
- XUT : Transcrits instables sensibles à xrn1, « Xrn1-sensitive unstable transcripts »
- SUT : Transcrits stables non annotés, « Stable Unnanotated transcripts »
- eCUT : Formes allongées des CUT, « Elongated-CUT »
- NUT : Transcrits allongés en l'absence de Nrd1, « Nrd1-Unterminated transcripts »
- RNA-seq : Séquençage d'ARN
- Net-Seq : Séquençage des ARN en cours de transcription, « Native elongative transcripts sequencing »
- FISH : Hybridation in situ de sondes fluorescentes, « Fluorescent *in situ* Hybridation »

Table des matières

INTRODUCTION 10

I. LA TRANSCRIPTION GENIQUE 12

1. INITIATION DE LA TRANSCRIPTION 12
2. ELONGATION DE LA TRANSCRIPTION 22
3. TERMINAISON DE LA TRANSCRIPTION 25

II. LA TRANSCRIPTION PERVASIVE 27

1. CONTROLE DE L'INITIATION DE LA TRANSCRIPTION PERVASIVE 28
2. CONTROLE NUCLEAIRE CONDUISANT A LA DEGRADATION DES TRANSCRITS CRYPTIQUES 31
3. SURVEILLANCE CYTOPLASMIQUE DES ARN : LA VOIE DU NONSENSE-MEDIATED MRNA DECAY (NMD) 33
4. MISE EN EVIDENCE DES TRANSCRITS PERVASIFS DANS LA CELLULE 35

III. ROLES DE LA TRANSCRIPTION PERVASIVE 38

1. FONCTIONNALITE EN *TRANS* 38
2. FONCTIONNALITE EN *CIS* 39

IV. PROJET DE THESE 46

RESULTATS 47

I. DESCRIPTION DES MECANISMES D'INTERFERENCE TRANSCRIPTIONELLE CHEZ *SACCHAROMYCES CEREVISIAE* 47

1. RESULTATS PRECEDENTS 47
2. DEVELOPPEMENT D'UN SYSTEME RAPPORTEUR POUR L'ETUDE DES MECANISMES D'INTERFERENCE TRANSCRIPTIONNELLE 50

II. ROLE DE PHENOMENES D'INTERFERENCE TRANSCRIPTIONNELLE LIES A LA TRANSCRIPTION ANTISENS DANS LA REGULATION NATURELLE DE L'EXPRESSION GENIQUE (MANUSCRIT DE L'ARTICLE) 57

DISCUSSIONS ET PERSPECTIVES 103

CONCLUSION 114

MATERIELS ET METHODES 115

BIBLIOGRAPHIE 121

Table des figures

Figure 1 : Structure de l'ARN polymérase II..... 12
Figure 2 : (A) Modifications biochimiques du CTD de la polymérase (Egloff & Murphy, 2008) (B) Gradient de chaque modification et facteurs recrutés au cours du cycle de transcription (Heidemann et al, 2013) 13
Figure 3 : Modèle de la structure du PIC..... 14
Figure 4 : Assemblage et activation du complexe de pré-initiation de la transcription..... 15
Figure 5 : Différents niveaux de compaction de l'ADN 18
Figure 6 : Organisation des nucléosomes au niveau des promoteurs 19
Figure 7 : Différentes marques de modification des histones..... 21
Figure 8 : Méthylations co-transcriptionnelles et effecteurs ciblés 25
Figure 9 : Voie principale de terminaison de la transcription génique..... 26
Figure 10 : Origines et contrôles de la transcription pervasive 30
Figure 11 : Voie de terminaison Nrd1-Nab3-Sen1 (Adaptée de Porrua & Libri, 2015) 32
Figure 12 : Structure de l'exosome nucléaire (Adaptée de Kilchert et al, 2016) 32
Figure 13 : Devenir cellulaire de la transcription pervasive..... 37
Figure 14 : Impact de la transcription antisens sur la chromatine dans les régions promotrices et codantes des gènes (Murray et al, 2015)..... 39
Figure 15: Distribution des gènes suivant l'effet de la délétion du facteur de modification de la chromatine sur le niveau d'ARNm (M norm = log2 (abondance Δfacteur /abondance)). 48
Figure 16 : Classement des gènes réprimés par interférence transcriptionnelle ("Régulés") selon le facteur de modification de la chromatine spécifiquement impliqué dans le mécanisme..... 49
Figure 17 : (A) Représentation des résultats de RNA-seq. (B) Northern Blot montrant la répression du gène CPS1 par la transcription antisens (Résultat non publié, Feurbach F., Malabat C. et Jacquier A.) 52
Figure 18 : Test de croissance différentielle en milieu liquide..... 52
Figure 19 : Evolution du pourcentage de cellules CUP1_s_CPS1 au cours du test de compétition de croissance entre une souche cup1Δ et la souche CUP1_s_CPS1 à 20µm de CuSO4..... 53

Figure 20 : Evolution du pourcentage de cellules CUP1_s_CPS1 au cours du test de compétition de croissance entre une souche CUP1_s_CPS1 et la souche CUP1_s et as_CPS1 à différentes concentrations de CuSO4 54
Figure 21 : Northern-Blot pour la détection des transcrits sens et antisens de CUP1, au sein du système rapporteur associé à différentes combinaisons de promoteurs..... 54
Figure 22 : Test de l'expression des gènes rapporteurs de la construction URA3-GFP-mCherry 55
Figure 23 : Test de l'expression sens et antisens du système rapporteur 56
Figure 24 : Récapitulatif des régulations déduites des expériences d'interruption NNS pour PET10, CLD1, SHH3, MOH1, et ARO10 (Les rectangles bleus et oranges correspondent respectivement aux promoteurs géniques et antisens) 108
Figure 25 : Représentation de l'évolution du niveau d'Antisens / d'ARNm et des P values associées (Wilcoxon apparié) (Adapté de Alcid & Tsukiyama, 2016)..... 112
Figure 26 : Evolution de la taille des lncARN et de leur taille de recouvrement avec les gènes associés 113
Figure 27 : Schéma des deux systèmes rapporteurs développés..... 116
Figure 28 : Schéma de la méthode de mutagénèse utilisée pour l'insertion du NNS 117

Introduction

L'expression des gènes se produit à partir de l'information génétique contenue dans une séquence d'ADN et aboutit à la synthèse de produits fonctionnels protéiques, à rôle catalytiques et/ou structuraux. Chez les eucaryotes, elle dépend de plusieurs processus essentiels : *(i)* La transcription génique dans le noyau pour produire un pré-ARN messager (pré-ARNm) *(ii)* La maturation de celui-ci en ARN messager (ARNm) par l'ajout de coiffe et de queue polyA ainsi que l'épissage des introns *(iii)* L'export de cet ARNm vers le cytoplasme *via* les pores nucléaires et jusqu'au lieu optimal de traduction *(iv)* Sa traduction en une chaîne peptidique par les ribosomes *(v)* L'assemblage en protéines et complexes protéiques avec une conformation précise.
Toutes ces étapes sont interconnectées et soumises à des contrôles-qualité multiples pour assurer une élimination d'intermédiaires ou de produits incorrects qui pourraient avoir un effet néfaste pour la cellule. La régulation de l'expression des gènes intervient à différents niveaux, pour permettre une adaptation plus ou moins rapide de la cellule face à des changements environnementaux. En particulier, les régulations transcriptionnelles pour la régulation des gènes sont très diverses, et permettent une régulation fine du niveau d'expression.

Chez les eucaryotes, la transcription est un processus élémentaire qui représente la première étape de l'expression des gènes. Elle requiert notamment l'action de trois complexes macromoléculaires appelés ARN polymérases (ARN pol), possédant une activité catalytique de transcription d'ARN à partir d'une matrice ADN. Ces trois polymérases partagent des unités protéiques communes, mais ont également des sous-unités spécifiques responsables de leur spécificité de produits ARN et de localisation. La polymérase I se situe dans le nucléole et intervient dans la biogénèse des ribosomes, plus particulièrement par la transcription du précurseur majoritaire des ARN ribosomiques (45S) (Han *et al*, 2017). La polymérase III, présente dans le noyau, synthétise principalement les précurseurs des ARN de transfert, mais transcrit également le précurseur d'ARN ribosomique 5S ainsi que divers autres petits ARN structuraux (Turowski *et al*, 2016). L'ARN polymérase II intervient dans la transcription des gènes en ARNm, et permet aussi la synthèse des petits ARN nucléaires (snARN) et nucléolaires (snoARN) (Matera *et al*, 2007).

Grâce à la mise au point de nouvelles techniques, notamment celles de séquençage massif, les études de transcriptomes ont pu s'affranchir du biais de l'annotation du génome. Ces nouvelles technologies ont permis d'observer que la transcription par l'ARN polymérase II n'est pas limitée à des unités fonctionnelles bien définies comme les gènes. Il y a en fait une transcription plus généralisée et "pléthorique" (l'anglicisme « pervasive » est souvent utilisé), souvent cachée (en cela qu'elle ne génère pas d'ARN suffisamment stables pour s'accumuler à des niveaux détectables) (Jacquier, 2009; Tisseur *et al*, 2011).

Chez les métazoaires, des études à l'échelle du génome, dans le cadre de projets comme le « RIKeN mouse Gene encyclopedia project » chez la souris ou « FANTOM3 » chez l'Homme, ont effectivement révélé que 80% à 98% du génome est activement transcrit. Ces études ont apporté la description de « nouveaux » transcrits non-codants, qui sont produits à partir de régions que l'on pensait transcriptionnellement silencieuses, ou en antisens des gènes (Jacquier, 2009).

Chez l'Homme, des transcrits pervasifs de types divers ont pu être décrits et associés à des régions déterminantes pour la transcription génique comme les promoteurs, les terminateurs, les « enhancers », ou les sites d'épissage (Jacquier, 2009; Tisseur *et al*, 2011). Leur étude a permis d'identifier des cas ponctuels où ces ARN non-codants seraient fonctionnels et régulateurs pour l'expression des gènes, même si la question de la fonctionnalité globale de la transcription pervasive subsiste.

C'est d'abord chez la levure de bière *Saccharomyces cerevisiae* que les ARNs pervasifs ont été mis en évidence (Wyers *et al*, 2005; van Dijk *et al*, 2011; Xu *et al*, 2009; Jacquier, 2009). Ce champignon unicellulaire, de complexité inférieure et totalement dépourvu des petits ARN interférents (siARNs, miARNs) ainsi que de protéine Argonaute, représente un organisme modèle d'intérêt pour l'étude du rôle potentiel de la transcription pervasive dans la régulation de l'expression des gènes. Son génome représente environ 12 Mb et est réparti sur 16 chromosomes nucléaires (et un chromosome mitochondrial), pour un total de 6604 séquences codantes identifiées. 78% de ces séquences correspondent à des gènes, 11% n'ont pas encore été caractérisées et 10% sont des séquences codantes supposées, dites « Dubious » (http://www.yeastgenome.org/genomesnapshot).

I. La transcription génique

1. Initiation de la transcription

L'initiation de la transcription représente un point de contrôle critique dans l'expression des gènes. Elle se caractérise d'abord par l'assemblage du complexe de Pré-Initiation (PIC) qui constitue la machinerie minimale requise pour une transcription basale des gènes, et interagit avec les co-activateurs et répresseurs transcriptionnels pour activer ou inhiber la transcription.

a. Assemblage du PIC

Cet assemblage est permis par un recrutement séquentiel des facteurs généraux de transcription (GTF) et de l'ARN polymérase II au niveau du promoteur génique.

Chez *Saccharomyces cerevisiae*, l'ARN polymérase II est un complexe macromoléculaire enzymatique constitué de 12 sous-unités nommées Rpb1 à Rpb12 (Figure 1). Dix sous-unités protéiques sont très conservées chez les eucaryotes et seules Rpb4 et Rpb9 sont non-essentielles pour la levure. Rpb1 et Rpb2 ont un rôle structural très important car elles comprennent le pore pour l'entrée des ribonucléotides, ainsi que le site de fixation à l'ADN.

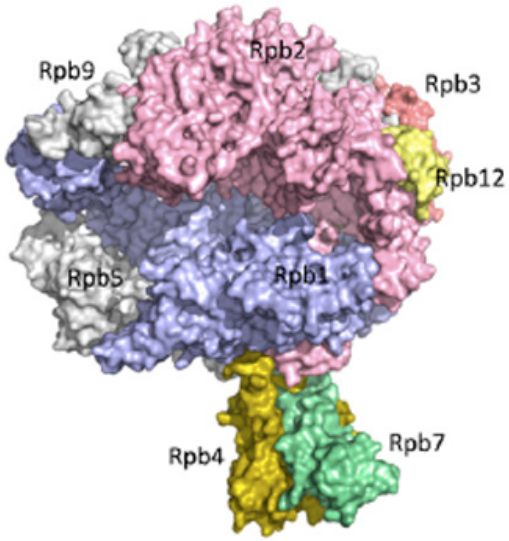


Figure 1 : Structure de l'ARN polymérase II
(Hahn,2004, Nature Structural and molecular biology)

D'autre part, Rpb1 est une sous-unité particulière, car son domaine C-terminal contient une série de répétition de l'hepta-peptide YSPTSTS (26 répétitions chez la levure et plus chez les organismes plus complexes) appelée CTD pour « Carboxy Terminal Domain », au rôle

majeur dans la régulation de la transcription. En effet, le CTD va subir des modifications biochimiques tout au long du cycle de transcription. Ces modifications peuvent être de type phosphorylation, isomérisation ou glycosylation. Elles vont cibler un acide aminé précis de l'hepta-peptide répété, et ont un rôle crucial dans le recrutement de complexes protéiques au cours de la transcription (Figure 2) (Hahn, 2004; Hahn & Young, 2011; Meinhart *et al*, 2005).

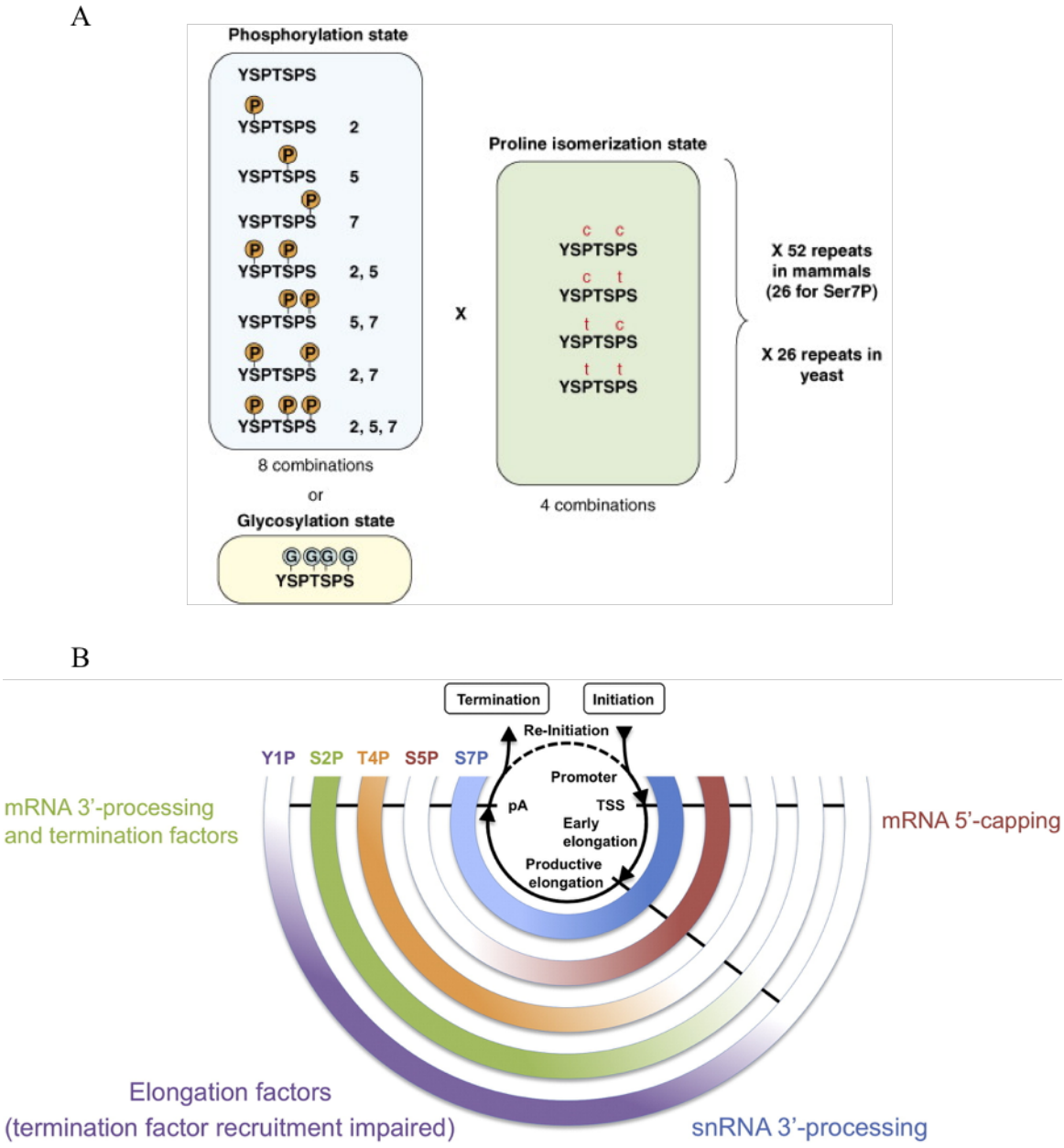


Figure 2 : (A) Modifications biochimiques du CTD de la polymérase (Egloff & Murphy, 2008) (B) Gradient de chaque modification et facteurs recrutés au cours du cycle de transcription (Heidemann et al, 2013)

Les GTF sont des complexes protéiques qui sont spécifiques à chaque type de polymérase. Pour la transcription génique, ils sont impliqués dans la reconnaissance de la séquence

d'ADN du promoteur, dans le recrutement et la stabilisation de l'ARN pol II sur le promoteur ainsi que dans le changement conformationnel nécessaire à son activation (Rhee & Pugh, 2012; Hahn, 2004; Sainsbury *et al*, 2015).

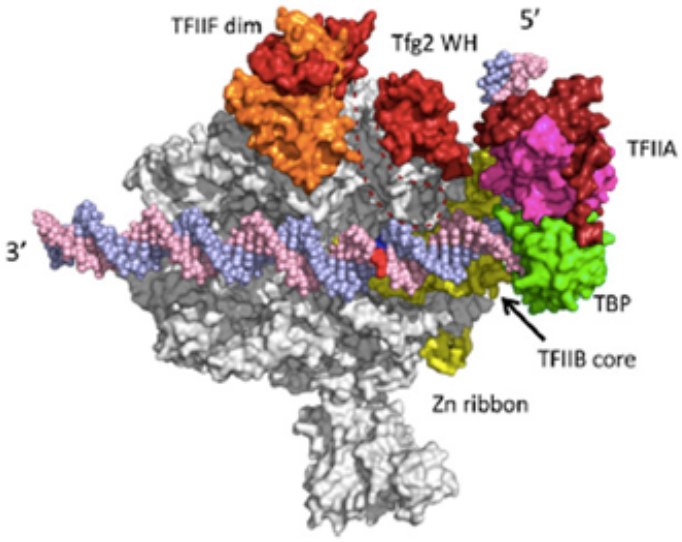


Figure 3 : Modèle de la structure du PIC
(Hahn,2004, Nature Structural and molecular biology)

L'assemblage du PIC débute avec le recrutement de la TBP pour « TATA-binding Protein » au niveau du promoteur. La TBP est recrutée au niveau de l'ADN par l'intermédiaire du complexe TFIID pour 90% des gènes dont le promoteur ne contient pas la séquence correspondant à l'élément TATA. Pour 10% des gènes, majoritairement associés à des réponses aux conditions de stress, et dont le promoteur contient l'élément TATA, le recrutement de la TBP fait appel au complexe SAGA pour « Spt-Ada-Gcn5 acetyltransferase ». L'élément TATA a une séquence consensus TATAWAWR et est situé environ 30 bases en amont du site de démarrage de transcription. Les complexes TFIID et SAGA sont tous deux constitués de facteurs associés à la protéine TBP, appelés TAF, qui peuvent interagir avec d'autres éléments de séquence que l'élément TATA, tels que l'élément initiateur (Inr) ou l'élément de reconnaissance par TFIIB (Bre) (Rhee & Pugh, 2012; Hahn, 2004; Sainsbury *et al*, 2015).

Le complexe contenant la protéine TBP interagit donc de manière spécifique ou non avec la séquence d'ADN du promoteur, et a un rôle de plateforme moléculaire pour l'assemblage des autres composants du PIC. Il permet aussi une distorsion de la séquence d'ADN, nécessaire pour la détection du site de démarrage de transcription (TSS). Le facteur TFIIA est recruté

avec la TBP, et assure un rôle stabilisateur pour sa fixation sur le promoteur (Rhee & Pugh, 2012; Hahn, 2004; Sainsbury *et al*, 2015).

Le facteur TFIIB entre ensuite dans la composition du complexe d'initiation. Il est essentiel au recrutement de l'ARN pol II, par son interaction directe avec la séquence d'ADN, la TBP, et le domaine catalytique de l'ARN pol II. La polymérase II est ensuite co-recrutée avec le facteur TFIIF. Ainsi, TFIIF permet le recrutement de l'ARN pol II, mais il joue également un rôle dans l'indentification du TSS. A ce stade, le cœur du PIC est formé et occupe 60 pb environ sur l'ADN, mais il est encore inactif (Figures 3 et 4).

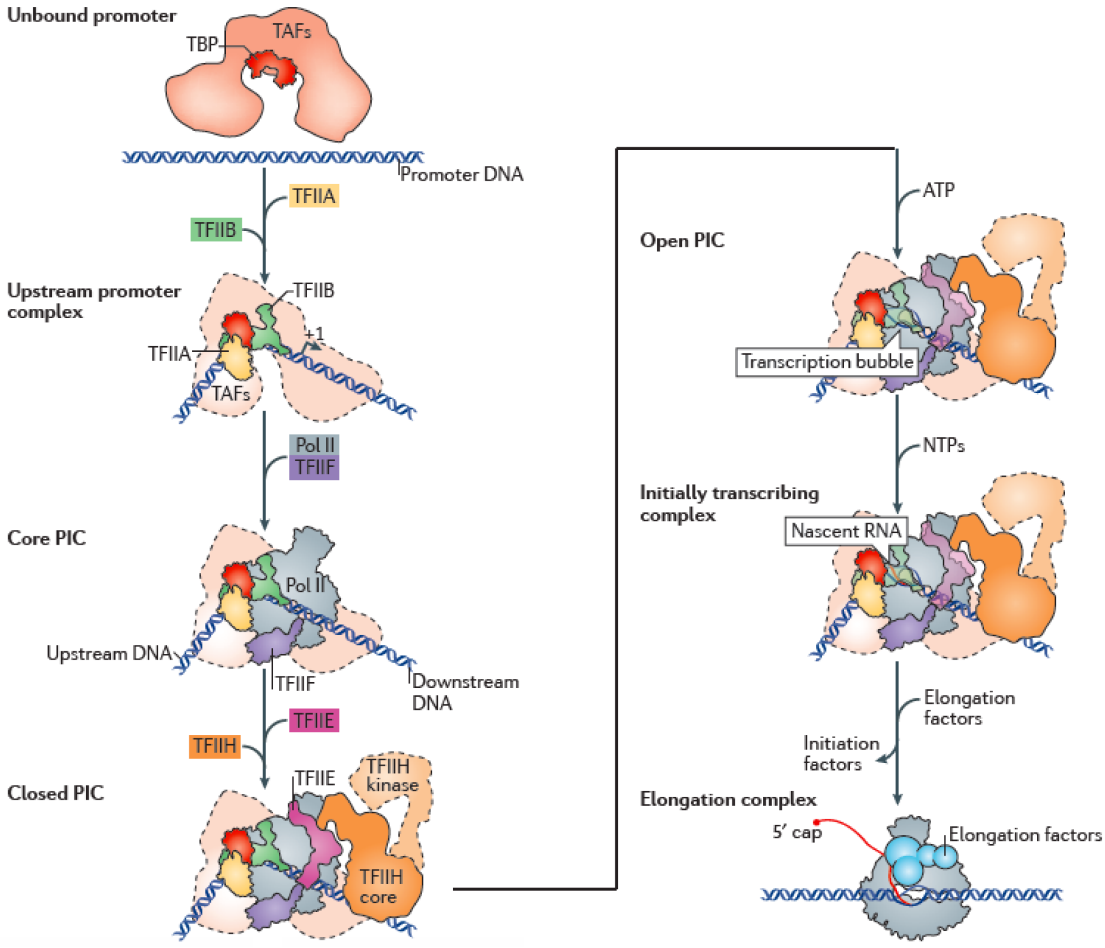


Figure 4 : Assemblage et activation du complexe de pré-initiation de la transcription (Adaptée de Sainsbury et.al., 2005, Nature Reviews)

b. Ouverture de l'ADN et démarrage de transcription

TFIIE et TFIIH sont recrutés au niveau du PIC, indépendamment de l'ARN pol II. Un changement conformationnel conduit à l'insertion de la séquence d'ADN dans la polymérase

II, ainsi qu'à une ouverture de l'ADN pour former la bulle de transcription. TFIIH participe à cette ouverture du PIC grâce à ses activités hélicases ATP-dépendantes portées par les protéines Xpb et SLL2, tandis que TFIIE stabilise l'ouverture (Rhee & Pugh, 2012; Hahn, 2004; Sainsbury *et al*, 2015).

Grâce à l'activité hélicase de TFIIH, la polymérase scanne la séquence d'ADN en partant du promoteur, afin de repérer le site de démarrage de transcription. La sous-unité Kin28 de TFIIH est également à l'origine d'une phosphorylation en sérine 5 (S5P) du CTD de l'ARN pol II, qui va être importante pour les étapes suivantes de la transcription car elle sera reconnue par différents facteurs. L'ARN pol II utilise ensuite comme substrat des ribonucléotides pour synthétiser le début de la chaîne d'ARN. A cet instant, l'hybride ADN-ARN formé est trop court pour être stable. Les brins matrices et codants peuvent avoir tendance à se ré-hybrider, ce qui peut générer des phénomènes d'initiation abortive avec libération de petits fragments d'ARN (Rhee & Pugh, 2012; Hahn, 2004; Sainsbury *et al*, 2015).

c. Fidélité de l'initiation de la transcription

La faible spécificité de sélection du TSS par le PIC conduit à la production d'isoformes des transcrits et peut avoir un rôle régulateur pour l'expression des gènes (Arribere & Gilbert, 2013; Malabat *et al*, 2015). En effet, le choix d'un site alternatif de démarrage de transcription peut avoir plusieurs conséquences :

- Le transcrit peut avoir conservé le site de démarrage de la traduction (codon AUG) mais avec une extrémité 5' non traduite différente (5' UTR). Dans ce cas l'efficacité de traduction et donc le niveau d'expression du gène sera impacté (Arribere & Gilbert, 2013).
- Le TSS alternatif peut avoir généré un autre codon AUG, dans la même phase de traduction mais en aval du premier, qui conduirait à la production d'une protéine tronquée (Arribere & Gilbert, 2013).
- Le TSS sélectionné peut être en amont de la phase ouverte de lecture du gène et créer un petite phase ouverte de lecture dite uORF qui fera du transcrit généré une cible pour la dégradation liée au contrôle qualité des ARNm (Malabat *et al*, 2015).
- Enfin, les TSS alternatifs peuvent être situés en amont ou en aval de sites de terminaison précoce (de type « NNS », voir plus loin) et ainsi conduire, selon les conditions de croissance, à une transcription abortive ou normale (Voir Partie I.3).

d. Architecture de la chromatine au niveau des promoteurs

L'information génétique de la cellule est codée dans le noyau, au niveau de l'ADN qui est lui-même empaqueté dans la chromatine au sein des chromosomes (Figure 5).

Ainsi, le premier niveau de compaction de l'ADN est son enroulement autour des octamères d'histones, pour former les structures appelées nucléosomes. Chez *S. cerevisiae*, les histones sont au nombre de 5 : H1, H2A, H2B, H3 et H4. L'histone H1 a un rôle très limité dans l'initiation de la transcription (il intervient plutôt dans la régulation de la compaction de la chromatine au cours de la sporulation), mais les quatre autres sont présents en deux copies au sein de la structure du nucléosome (Rando & Winston, 2012). Les protéines d'histones (de 11 à 23 kDa) ont la particularité d'avoir un domaine de repliement nommé « histone fold » à l'extrémité C-terminale, indispensable à l'interaction entre histones pour former la structure du nucléosome. Ces protéines ont également une queue N-terminale, exposée à l'extérieur du nucléosome et qui sont la cible de nombreuses modifications post-traductionnelles (Rando & Winston, 2012) (Voir paragraphe 3).

Pour initier la transcription, l'ARN polymérase II et les facteurs généraux de transcription doivent accéder à l'ADN empaqueté dans la chromatine. Les nucléosomes constituent donc une première barrière physique à l'assemblage de la machinerie de transcription au niveau du promoteur. Les modifications de la chromatine ont alors un fort impact sur la transcription des gènes et les promoteurs géniques présentent une architecture bien particulière de leur chromatine (Hahn & Young, 2011; Rando & Winston, 2012).

Les promoteurs géniques ont une organisation de la chromatine bien spécifique, qui diffère selon leur niveau et fréquence d'expression. Tous les promoteurs géniques se caractérisent par une région pauvre en nucléosome, appelée NFR pour « Nucleosome Free Region », qui est nécessaire à l'assemblage du PIC. Cette NFR est entourée de deux nucléosomes spécifiques, le nucléosome situé juste en amont (-1) étant la cible des régulations transcriptionnelles, et le nucléosome situé au niveau du TSS (+1) qui doit être décalé pour l'initiation transcriptionnelle (Rando & Winston, 2012; Cairns, 2009; Venters & Pugh, 2008).

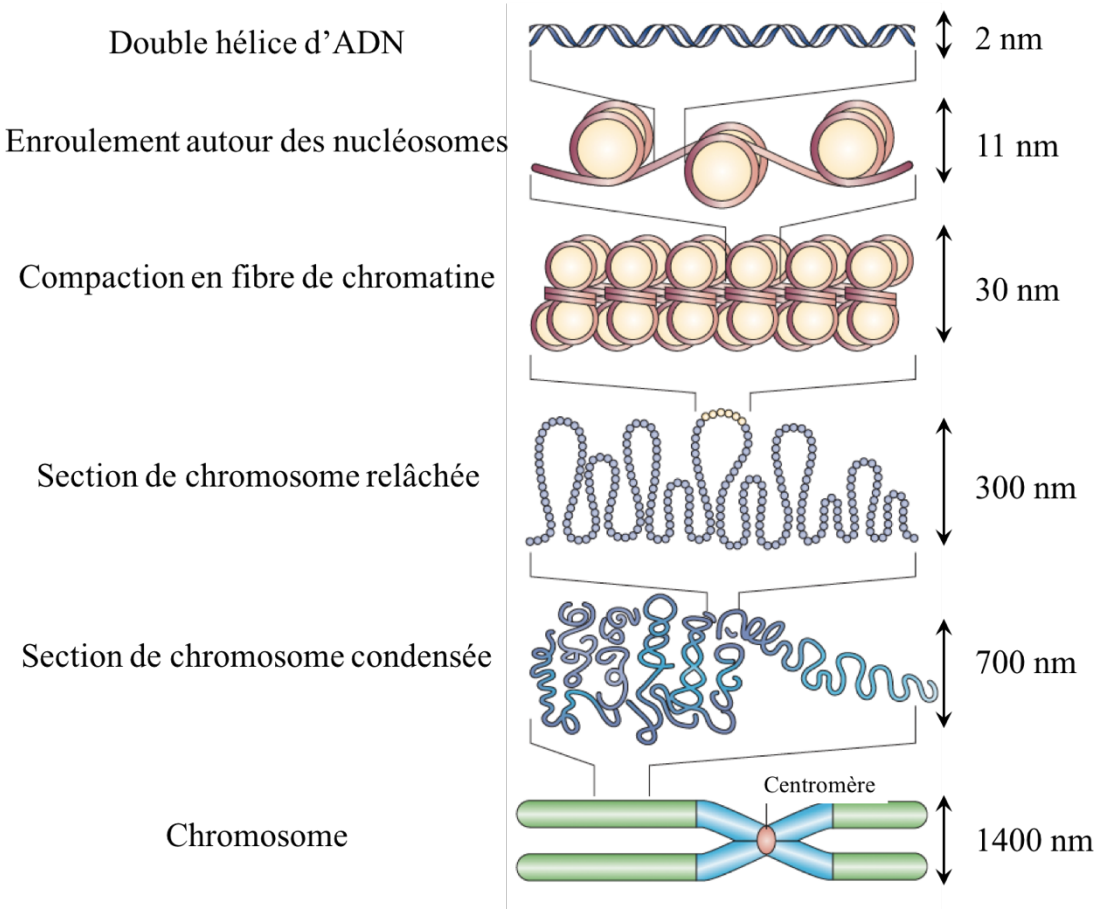


Figure 5 : Différents niveaux de compaction de l'ADN
(Adaptée de Schlissel, 2003)

Les promoteurs présentent généralement le variant d'histone H2AZ. Ce variant est présent sur 2/3 des gènes de *S. cerevisiae*, principalement au niveau du nucléosome +1, et dans une moindre mesure sur le nucléosome -1. Son incorporation fait intervenir le complexe SWR1 qui catalyse l'échange de dimères H2A-H2B contre des dimères H2AZ-H2B en présence d'ATP. Bien qu'il soit présent au niveau d'une grande partie des gènes, ce variant d'histone a un rôle qui n'est pas encore suffisamment décrit (Rando & Winston, 2012; Cairns, 2009; Venters & Pugh, 2008).

Les gènes « domestiques » qui sont impliqués dans les fonctions basiques pour la croissance cellulaire sont plutôt régulés par TFIID que par SAGA et sont dépourvus d'élément TATA. Ils ont peu de bruit transcriptionnel car ils ont une expression qui est dite constitutive. Ces gènes sont peu affectés par la délétion de gènes régulateurs de la chromatine et leurs promoteurs présentent une large NFR (Rando & Winston, 2012) (Figure 6).

Les gènes de stress sont eux majoritairement régulés par SAGA et possèdent un élément TATA. Ces gènes se caractérisent par une expression bruitée ou dite en « burst ». En effet, ils sont généralement très peu exprimés, mais peuvent être fortement et transitoirement induits dans certaines conditions environnementales. Ces gènes sont régulés par un large panel de facteurs de remodelage de la chromatine et ont une architecture très variable au niveau du promoteur, avec une NFR moins bien définie que pour les gènes domestiques (Figure 6). La fixation des facteurs de transcription y est gênée par les nucléosomes, mais l'échange rapide de nucléosomes permet une accessibilité transitoire pour la machinerie d'initiation de la transcription. La compétition entre les nucléosomes et la transcription participe ainsi à la variabilité intercellulaire que l'on appelle « bruit transcriptionnel » (Rando & Winston, 2012).

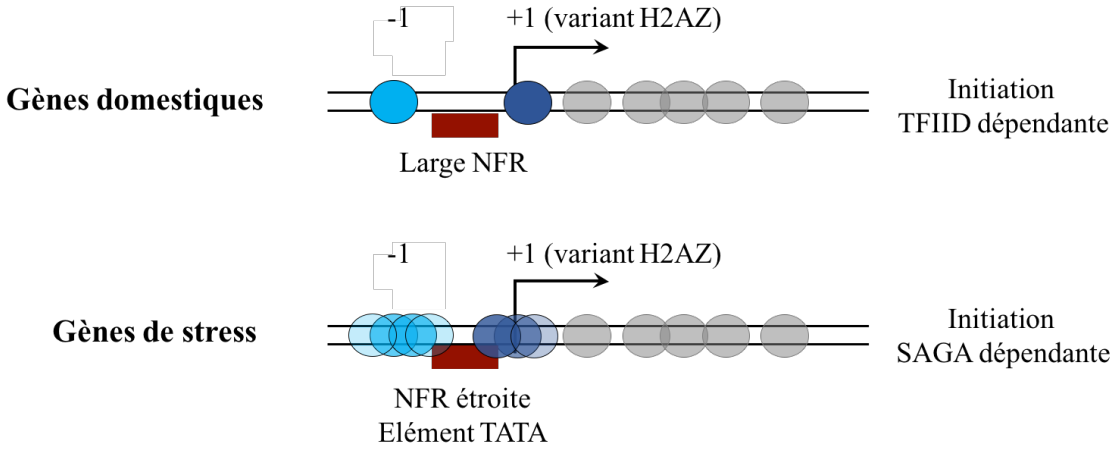


Figure 6 : Organisation des nucléosomes au niveau des promoteurs
(Adaptée de Rando & Winston, 2012)

Des changements transcriptionnels conduisent à une modification de la structure chromatinienne des promoteurs. Généralement, un haut niveau de transcription conduit à une expulsion du nucléosome -1 et à une extension de la NFR.

La structure chromatinienne des promoteurs constitue un élément clé dans la régulation transcriptionnelle des gènes afin que la cellule puisse adapter son programme d'expression au conditions environnementales (Rando & Winston, 2012).

e. Régulations transcriptionnelles à l'étape d'initiation

On oppose la transcription basale, de faible niveau et nécessitant la machinerie d'initiation minimale, à la transcription activée faisant intervenir des régulateurs transcriptionnels divers. La régulation de la transcription a majoritairement lieu au niveau de l'initiation de la transcription, et est réalisée de différentes façons :

- Par le recrutement de co-activateurs transcriptionnels et de GTF au niveau des promoteurs : il a été observé que l'activation de l'expression des gènes est corrélée à une augmentation du niveau de GTF et de facteurs de transcription sur les régions promotrices ainsi que sur les séquences régulatrices UAS pour « Upstream Activating Sequences » (sites ciblés par les facteurs de transcription, en amont du PIC) (Hahn & Young, 2011).

- Par des changements conformationnels de la machinerie de transcription pour augmenter son activité : ce type de régulation s'effectue par l'intermédiaire du complexe Médiateur. Ce complexe de 25 sous-unités protéiques joue un rôle d'intermédiaire entre les régulateurs transcriptionnels et les GTF. Il permet effectivement une intégration de divers signaux de facteurs de transcription, qui se traduit par un changement conformationnel et a des conséquences directes sur la formation du complexe de pré-initiation (stimulation du recrutement de l'ARN pol II, stabilisation du PIC…) (Hahn & Young, 2011).

- La modification de la structure de la chromatine : la modification de la chromatine peut avoir lieu par des modifications chimiques au niveau des queues d'histones composant le nucléosome. Ces modifications chimiques peuvent être des méthylations, acétylations, phosphorylations et ubiquitinylations des queues d'histones. Elles sont apposées ou retirées par des enzymes spécifiques, qui peuvent être recrutées par des facteurs de transcription (Hahn & Young, 2011; Rando & Winston, 2012; Millar & Grunstein, 2006).

  L'acétylation a globalement un impact positif pour la transcription, puisqu'elle neutralise les charges sur les queues d'histones et diminue ainsi leur interaction avec l'ADN au sein des nucléosomes. Cependant, l'ensemble des marques de modification de la chromatine constitue un code histone complexe, qui peut avoir un effet négatif ou positif pour l'expression génique (Rando & Winston, 2012).

  A côté des modifications d'histone, des remodelages de la chromatine sont aussi régulateurs pour l'initiation de la transcription. Ils sont permis par des complexes de remodelage dont l'activité dépend de la présence d'ATP et qui peuvent soit expulser des histones, soit les faire glisser d'une position à une autre.

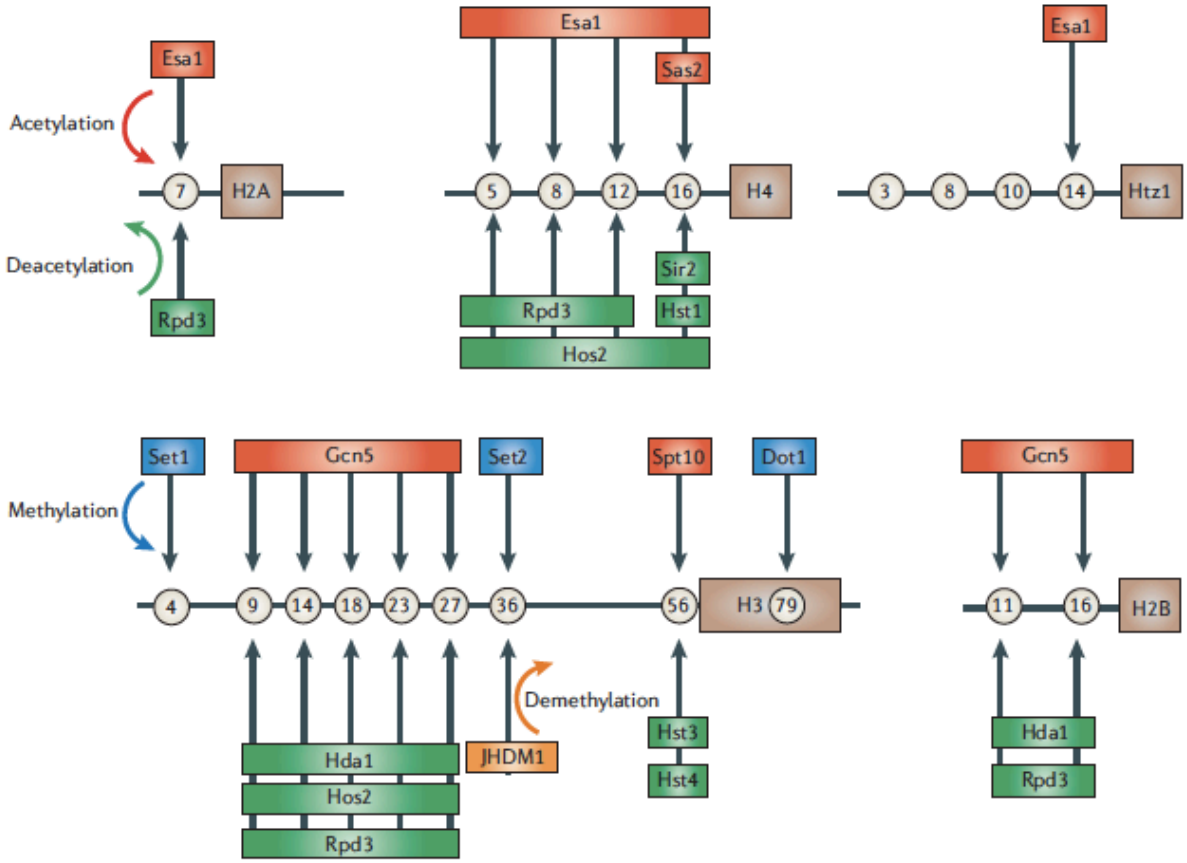


Figure 7 : Différentes marques de modification des histones
(Millar et. al., 2006, Nature Reviews Molecular Cell Biology)

Le complexe SWI/SNF, par exemple, altère la structure des nucléosomes pour permettre la fixation d'activateurs transcriptionnels et donc une levée de répression transcriptionnelle. Ce complexe agit en combinaison avec d'autre régulateurs transcriptionnels, son recrutement au promoteur est permis par une interaction directe avec des activateurs transcriptionnels, puis son association est stabilisée par son bromo-domaine (domaine lié aux queues d'histone acétylées) grâce à la reconnaissance des marques d'acétylation déposées par Gcn5 (Hahn & Young, 2011; Rando & Winston, 2012).

Le remodeleur RSC est un complexe essentiel très homologue à SWI/SNF, qui expulse les nucléosomes pour le maintien des NFR au niveau de nombreux promoteurs. Les remodeleurs ISW (Isw 1 et Isw 2) font plutôt glisser les nucléosomes d'une position à une autre. Isw2 par exemple, favorise le déplacement des

nucléosomes vers les NFR ou les TSS, et a un donc un rôle d'occlusion des promoteurs qui est négatif pour l'initiation de la transcription (Hahn & Young, 2011; Rando & Winston, 2012).

2. Elongation de la transcription

a. Mécanisme d'élongation transcriptionnelle

Pour la transition entre l'initiation de la transcription et son élongation, le complexe de transcription doit quitter la région promotrice. Ce phénomène s'appelle le dégagement du promoteur (« promoter clearance »). Dans les premiers temps de l'élongation, la bulle de transcription formée en fin d'initiation reste à sa position initiale et s'étend en taille au fur et à mesure de l'extension de l'ARN transcrit. Lorsque le transcrit a atteint la taille de 17 bases environ, l'extrémité 5' de l'ARN se dés-hybride de l'ADN matrice et pourra ensuite entrer dans le tunnel de sortie de la polymérase II. L'hybride ARN-ADN aura finalement une longueur d'environ 8 nucléotides (Hahn, 2004; Luse, 2013).

Pour le dégagement du promoteur, l'ARN pol II doit par ailleurs perdre ses interactions avec les facteurs d'initiation de la transcription et l'ADN du promoteur. Cette transition est favorisée par la marque S5P du CTD. Cette phosphorylation va également permettre de recruter les facteurs nécessaires à l'ajout de la coiffe (guanosine modifiée et méthylée ajoutée en 5') sur l'ARN naissant, afin d'augmenter sa stabilité et permettre le recrutement d'autres facteurs. Après la synthèse d'environ 30 bases d'ARN, l'ARN pol II quitte le PIC et le promoteur. Les GTF et co-activateurs transcriptionnels restent pour la plupart associés à la région promotrice constituant un échafaudage qui permet une ré-initiation rapide de la transcription. Seuls les facteurs TFIIB et TFIIH devront être recrutés à nouveau avec l'ARN pol II (Hahn, 2004; Luse, 2013).
Au début de la phase d'élongation, la polymérase II fait une pause et s'accumule en aval du TSS. Cette pause joue un rôle clé car elle permet d'effectuer un contrôle-qualité du transcrit (notamment de l'ajout de sa coiffe) avant de passer à l'élongation productive (Jonkers & Lis, 2015).

Il a été montré que même si la polymérase II est très processive *in vitro* sur de l'ADN nu, elle n'est pas très efficace pour transcrire une matrice ADN empaquetée dans la chromatine. Plusieurs facteurs d'élongation (EF) vont donc être recrutés au niveau du complexe de transcription pour faciliter le processus dynamique de l'élongation et en réguler sa vitesse.

Etant modulaires, les EF vont pouvoir interagir à la fois entre eux, avec le CTD de la polymérase et avec l'ARN naissant (Battaglia *et al*, 2017; Jonkers & Lis, 2015). L'état de phosphorylation du CTD joue un rôle clé dans leur recrutement. Les kinases Ctk1 et Bur1 sont en effet recrutées par la phosphorylation S5P du CTD (cette modification va être progressivement enlevée par la phosphatase Ssu72 au cours de l'élongation). Ces kinases vont pouvoir à leur tour phosphoryler la sérine 2 du CTD ainsi que la protéine Spt5 comprise dans le complexe DSIF (en association avec Spt4) pour permettre leur fonction de plate-forme moléculaire pour le recrutement d'autres EF. Parmi ces EF, on compte notamment les complexes Paf, FACT et le facteur d'élongation Spt6. D'autre part, DSIF va aider à prévenir les pauses de l'ARN pol II en cours d'élongation, en coopération avec le facteur d'élongation TFIIS qui va intervenir en cas de blocage de l'enzyme sur sa matrice (Battaglia *et al*, 2017; Meinhart *et al*, 2005; Jonkers & Lis, 2015; Hahn, 2004). Les EF sont interdépendants, et coopèrent dans le but ultime de permettre la progression de la polymérase à travers la barrière physique représentée par la chromatine, ainsi que le repositionnement des nucléosomes pour refermer la chromatine après son passage. Pour cela des modifications de la chromatine vont être nécessaires.

b. Modifications co-transcriptionnelles de la chromatine

La polymérase en cours d'élongation doit déplacer les nucléosomes pour accéder à la matrice d'ADN. Les gènes fortement transcrits seront d'ailleurs caractérisés par un espacement inter-nucléosomal plus petit sur la région codante. Si la transcription est très forte, on observera même une diminution d'occupation des nucléosomes sur la région codante (Rando & Winston, 2012). Ces remodelages de la chromatine sont permis par la coopération de plusieurs facteurs de modification de la chromatine, dont le recrutement est orchestré par les modifications du CTD de la polymérase II.

Rôle complexe de la méthylation par Set1 apposée en cours d'élongation

Set1 est une histone-méthyl transférase appartenant au complexe COMPASS qui est recruté au cours de la transcription. Set1 interagit directement avec la polymérase par la phosphorylation en sérine 5 du CTD, mais elle possède également un domaine de reconnaissance à des motifs sur l'ARN. D'autre part, son action est dépendante de la présence du facteur d'élongation Paf et d'une mono-ubiquitinylation de la lysine 123 de l'histone H2B. Le profil de méthylation généré par Set1 varie suivant la région transcrite avec un gradient de méthylation H3K4 le long des gènes activement transcrits. En effet, la région 5' en amont du TSS correspondant au moment de l'initiation de la transcription, est ciblée par une tri-

méthylation H3K4. Cette modification est une marque de la transcription active et est notamment reconnue par le domaine PHD des complexes Histone-acétyl transférase (HAT) Nua3, Nua4, Hbo1 et SAGA. Ces complexes régulent positivement la transcription génique car ils permettent une acétylation de la chromatine au niveau des promoteurs. En revanche, les di-méthylations seront plutôt présentes en 5' mais en aval du TSS (correspondant à l'élongation précoce) et sont lues par le chromo-domaine de la protéine Set3 comprise dans le complexe Set3C également constitué des deux histone-déacétylases (HDAC) Hos1 et Hst1. Ce complexe a à la fois un rôle positif et négatif pour la transcription génique. En effet, il a été montré qu'il permet de contrôler la cinétique d'induction des gènes (Kim *et al*, 2012), mais il est généralement connu pour son effet répresseur sur la transcription (van Werven *et al*, 2012). Les mono-méthylations sont plutôt distribuées dans le corps de la région transcrite (Woo *et al*, 2017; Rando & Winston, 2012; Millar & Grunstein, 2006).

Coopération de l'HMT Set2 avec les facteurs d'élongation pour refermer la chromatine après le passage de la polymérase

Set2 est une HMT H3K36 qui se lie aux marques S5P et S2P du CTD. Elle est à l'origine d'une tri-méthylation H3K36me3 sur le corps des gènes activement transcrits. Sa fonction principale est le recrutement du complexe Rpd3S par le chromo-domaine d'un de ses constituants Eaf3, conduisant ainsi à une déacetylation par l'HDAC Rpd3. La marque H3K36me3 est également liée par le complexe ISW1. Ces deux complexes sont à l'origine d'une déacetylation et d'un remodelage, importants après le passage de la polymérase II pour refermer la structure de la chromatine et la rendre incompétente pour une initiation transcriptionnelle (Woo *et al*, 2017; Rando & Winston, 2012; Millar & Grunstein, 2006).
Ea3f entre également dans la composition du complexe NuA4, et deux composants de NuA3 (Nto1 et Pdp3) reconnaissent également la méthylation H3K36me3. Cela suggère que Set2 pourrait aussi avoir un rôle opposé, pour le recrutement d'HAT pendant la transcription (Woo *et al*, 2017).

Remodelage de la chromatine pendant l'élongation transcriptionnelle

Les chaperonnes d'histone Spt16 et Spt6 jouent un rôle clé dans l'ouverture de la structure de la chromatine pour la transcription. Spt6 contient un domaine de liaison à l'ADN et deux domaines d'interaction avec le CTD phosphorylé en sérine 2 de la polymérase qui permettent son recrutement en cours d'élongation. Son rôle est de réguler l'abondance des nucléosomes sur les régions fortement transcrites (son effet sur la transcription génique peut être positif ou

négatif). Elle est aussi requise pour la méthylation H3K36 (par l'HMT Set2) et pour la maintenance de la phosphorylation en sérine 2 du CTD de la polymérase II. Spt16 entre dans la composition du complexe FACT (avec la protéine Pob3) qui interagit avec le facteur d'élongation Paf pour diminuer les interactions entre l'ADN et les histones et retirer les dimères d'histones H2A-H2B afin de permettre le passage de la polymérase.

Le complexe RSC, en plus de son rôle dans le maintien des NFR pour l'initiation de la transcription, permet le remodelage de la structure des nucléosomes afin de permettre le passage de la polymérase en cours d'élongation. Son recrutement sur la chromatine est permis par son bromo-domaine, qui reconnait spécifiquement les acétylations d'histones déposées par le complexe SAGA et Nua4 (eux-mêmes recrutés par Set1) (Battaglia *et al*, 2017; Rando & Winston, 2012; Jonkers & Lis, 2015).

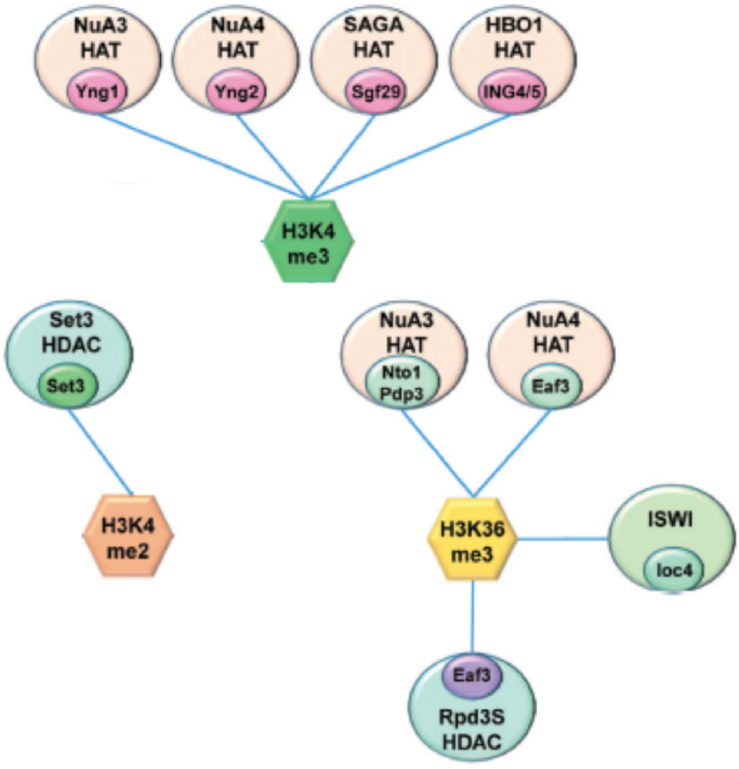


Figure 8 : Méthylations co-transcriptionnelles et effecteurs ciblés
(Adaptée de Woo et.al., 2017, Experimental & Molecular Medicine)

3. Terminaison de la transcription

La terminaison de la transcription se traduit par l'arrêt de la synthèse d'ARN, ainsi que la libération de l'ARN naissant et la dissociation de l'ARN polymérase II de la matrice d'ADN. Le mécanisme de terminaison dépend de l'intervention de facteurs de terminaisons

appartenant au facteur de clivage et de polyadénylation CPF, et au facteur de clivage comprenant les complexes CFIA, et CFIB (Porrua & Libri, 2015; Libri, 2015; Kuehner *et al*, 2011).

Après la transcription d'un site de polyadénylation, la reconnaissance d'éléments de séquences dans la région 3' non traduite de l'ARN (3'UTR) par les facteurs de terminaison conduit à une pause de la polymérase. Ces facteurs sont recrutés en cours d'élongation par la polymérase. La marque S2P du CTD est reconnue par la protéine Pcf11 du complexe CPF (*via* son domaine d'interaction avec le CTD, appelé CID), tandis que Clp1 du complexe CFIA interagit avec la sous-unité Rbp2. Il y a ensuite un clivage endonucléolytique de l'ARN naissant réalisé au niveau du site de polyadénylation par l'endonucléase Ysh1 (au sein de CPF). La coupure produit deux fragments d'ARN, le fragment en amont étant polyadénylé à son extrêmité 3' grâce à la poly-A polymérase Pap1, alors que le deuxième fragment est dégradé de 5' en 3' par l'exonucléase Rat1 (Porrua & Libri, 2015; Libri, 2015; Kuehner *et al*, 2011).

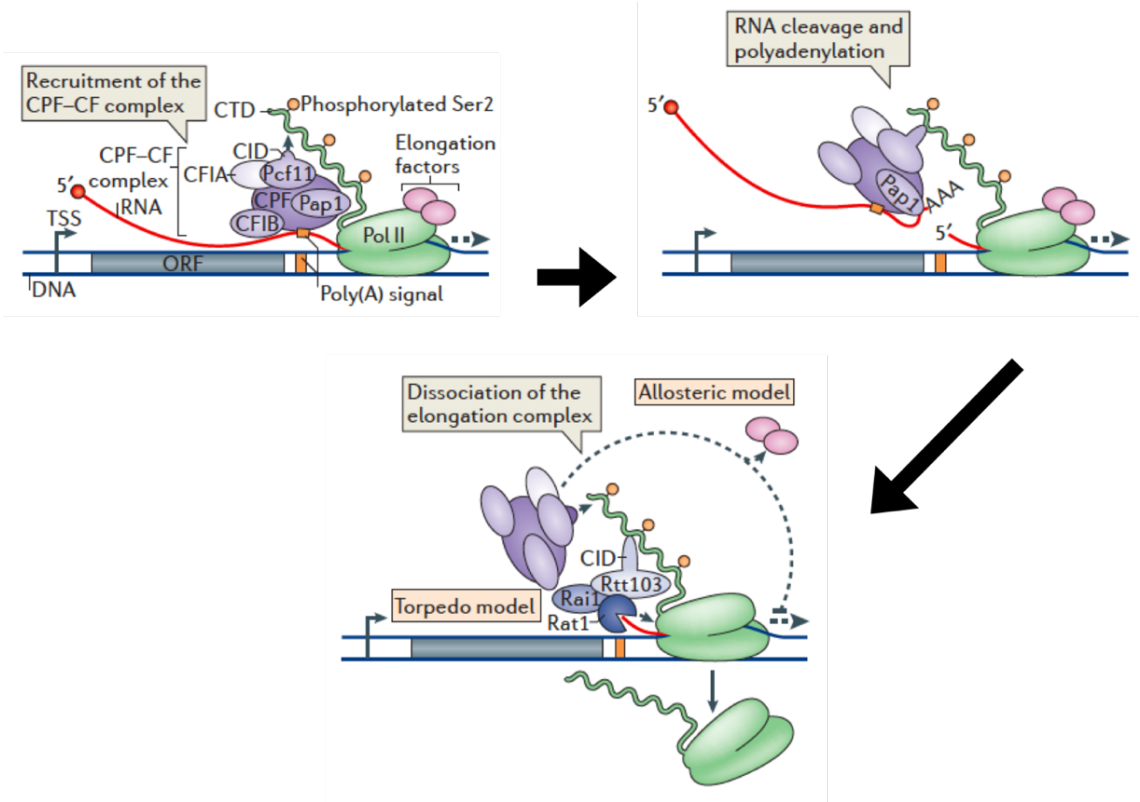


Figure 9 : Voie principale de terminaison de la transcription génique
(Adaptée de Porrua et.al., 2015, Nature Reviews Molecular Cell Biology)

Deux modèles coexistent pour expliquer la dissociation du complexe de transcription de sa matrice. Le premier est le modèle allostérique, dans lequel un changement conformationnel du complexe d'élongation après la fixation des facteurs de terminaison conduit à la perte des facteurs d'élongation, puis à la libération de l'ARN pol II de la matrice ADN. Le deuxième modèle est le modèle Torpedo, selon lequel le clivage endonucléolytique par Ysh1, et la dégradation exonucléolytique de l'ARN en aval de la coupure par Rat1, permettent une déstabilisation du complexe de transcription qui conduirait à sa libération (Porrua & Libri, 2015; Libri, 2015).

La terminaison de la transcription est également sujette à des régulations. En effet, différents sites de polyadénylation peuvent être reconnus pour un même transcrit. D'autre part, nous verrons par la suite qu'il existe une autre voie de terminaison caractéristique des transcrits non-codants qui peut être utilisée à la place de la voie classique CPF-CF pour réguler le niveau d'expression de certains gènes (Porrua & Libri, 2015).

II. La transcription pervasive

La transcription pervasive chez *Saccharomyces cerevisiae* a été mise en évidence il y a une dizaine d'années (Wyers *et al*, 2005). Elle est permise par l'ARN polymérase II et correspond à une transcription non-codante pléthorique, initiée à partir de régions d'ADN que l'on pensait transcriptionnellement silencieuses comme les promoteurs, les terminateurs et dans une moindre mesure à partir de régions intragéniques (Tisseur *et al*, 2011; Jacquier, 2009). Différentes études à grande échelle ont montré qu'une fraction substantielle de la transcription pervasive provenait de la bidirectionnalité des promoteurs géniques (Neil *et al*, 2009; Xu *et al*, 2009) qui conduit à la production de transcrits non-codants divergents par rapport à l'ARNm.

La cellule a développé des mécanismes de contrôle à plusieurs niveaux de l'expression génique pour contrer la transcription pervasive généralisée, ce qui fait que cette transcription de faible intensité est dite « cachée » ou cryptique (Jensen *et al*, 2013), puisque les transcrits qu'elle génère sont souvent trop instables pour être facilement détectables dans une souche sauvage. Les transcrits pervasifs ont d'ailleurs été classés en différentes catégories et nommés selon le mécanisme de contrôle cellulaire qui a été invalidé pour leur détection (Wyers *et al*, 2005; Xu *et al*, 2009; van Dijk *et al*, 2011) (Voir paragraphe 4). Ici, nous ferons surtout référence à la transcription pervasive de façon globale sans distinction de catégorie puisque

ces différentes catégories de transcrits ne sont pas clairement délimitées. En effet, selon les contrôles-qualité auxquels un même transcrit pervasif pourra partiellement échapper, il pourra répondre simultanément à des dénominations différentes.
Les transcrits pervasifs sont, comme les ARNm, des transcrits de l'ARN pol II coiffés et polyadénylés. Par ailleurs, de même que pour la transcription génique, la polymérase II engagée dans la transcription pervasive doit franchir la barrière de la chromatine pour accéder à la matrice ADN. Les transcrits générés subissent ensuite des contrôles-qualité post-transcriptionnels qui conditionneront leur devenir cellulaire.

1. Contrôle de l'initiation de la transcription pervasive

Les transcrits pervasifs ne correspondent pas à un bruit transcriptionnel aléatoire, car ils constituent des unités de transcription bien définies, produites à partir de régions précises (Neil *et al*, 2009). L'association du variant d'histone H2AZ avec la transcription pervasive (Gu *et al*, 2015), l'existence de TATA box cryptiques (Uwimana *et al*, 2017), l'induction spécifique de certains ARN pervasifs dans des conditions de croissance particulières, (Lardenois *et al*, 2011) et parfois leur régulation par des facteurs de transcription (van Werven *et al*, 2012; Nadal-Ribelles *et al*, 2014) suggèrent que ces transcrits pourraient au moins en partie être activement régulés.

La transcription pervasive est préférentiellement initiée à partir de régions pauvres en nucléosomes en 5' ou en 3' des gènes (Neil *et al*, 2009; Xu *et al*, 2009; Jacquier, 2009; Jensen *et al*, 2013). En particulier, une source importante de transcrits pervasifs est la bidirectionnalité intrinsèque des promoteurs géniques qui peuvent donc être le lieu d'assemblage de PIC distincts, initiateurs de transcription non-codante (Figure 8A). Dans ce cas, les contrôles de la transcription pervasive vont renforcer la directionnalité des promoteurs géniques pour éviter une initiation cryptique divergente. Plus particulièrement, il a été proposé que des phénomènes de boucles ou « gene looping » au niveau des gènes transcrits permettent de faire communiquer les régions d'initiation et de terminaison transcriptionnelles, afin d'assurer la directionnalité de la polymérase recrutée sur les promoteurs. Cette communication a lieu grâce à une interaction entre les facteurs de transcription recrutés sur les promoteurs et les facteurs de terminaison, notamment le facteur Ssu72 (Tan-Wong *et al*, 2012). Par ailleurs, il a été montré que le complexe d'assemblage de la chromatine CAF-I agit avec Rtt106 pour augmenter la densité en nucléosomes au niveau du nucléosome -1. Puis il y a dé-acétylation des histones incorporées par les HDAC Hst3 et Hst4. La combinaison de ces

deux mécanismes permet de limiter la transcription non-codante divergente (Marquardt *et al*, 2014). Le facteur de remodelage Isw2 permet également le repositionnement de nucléosomes près des promoteurs pour favoriser la directionnalité de la transcription (Whitehouse *et al*, 2007).

Des transcrits pervasifs non divergents peuvent aussi être générés à partir des promoteurs géniques. Ils sont généralement initiés en amont des ARNm et dans la même direction. Dans ce cas, les formations du PIC pervasif et du PIC génique seront en compétition pour l'utilisation des GTF disponibles (Neil *et al*, 2009; Jacquier, 2009). Bien que la transcription de l'ARNm soit plus robuste que la transcription non-codante, l'équilibre entre les deux PIC peut être déplacé en faveur de l'expression génique grâce à des signaux activateurs forts (Figure 8B).

Compte tenu de la disposition aléatoire, de la densité des gènes chez *S. cerevisiae* et du fait que la transcription divergente est un événement fréquent, cette dernière génère fréquemment des transcrits pervasifs en 3' du gène situé en amont lorsque celui-ci est orienté en tandem. De plus, il a été montré qu'il existe des évènements d'initiation pervasive en 3' qui sont indépendants du promoteur génique situé à proximité (Murray *et al*, 2012). La répression de ces initiations en 3' fait intervenir la voie Set2-Rpd3S pour rendre la chromatine répressive pour l'initiation cryptique par une dé-acétylation des Histones H4 (Figure 8C) (Churchman & Weissman, 2011).

La même voie Set2-Rpd3 permet également d'empêcher les phénomènes multiples d'initiations intragéniques cryptique, à l'occasion de l'ouverture de la chromatine pour la transcription génique (Figure 8 encadré D) (Malabat *et al*, 2015; Venkatesh *et al*, 2016). Cette voie supprime l'incorporation d'histones acétylés (par l'HAT Rtt109) sur les régions transcrites, en bloquant l'action de la chaperone d'histone Asf1 et en recrutant des remodeleurs de la chromatine Chd1 et Isw1b (Venkatesh & Workman, 2013; Smolle *et al*, 2012). Le facteur d'élongation Spt6 est également impliqué dans le ré-établissement de la structure de la chromatine après le passage de la polymérase II en élongation (Kaplan *et al*, 2003).

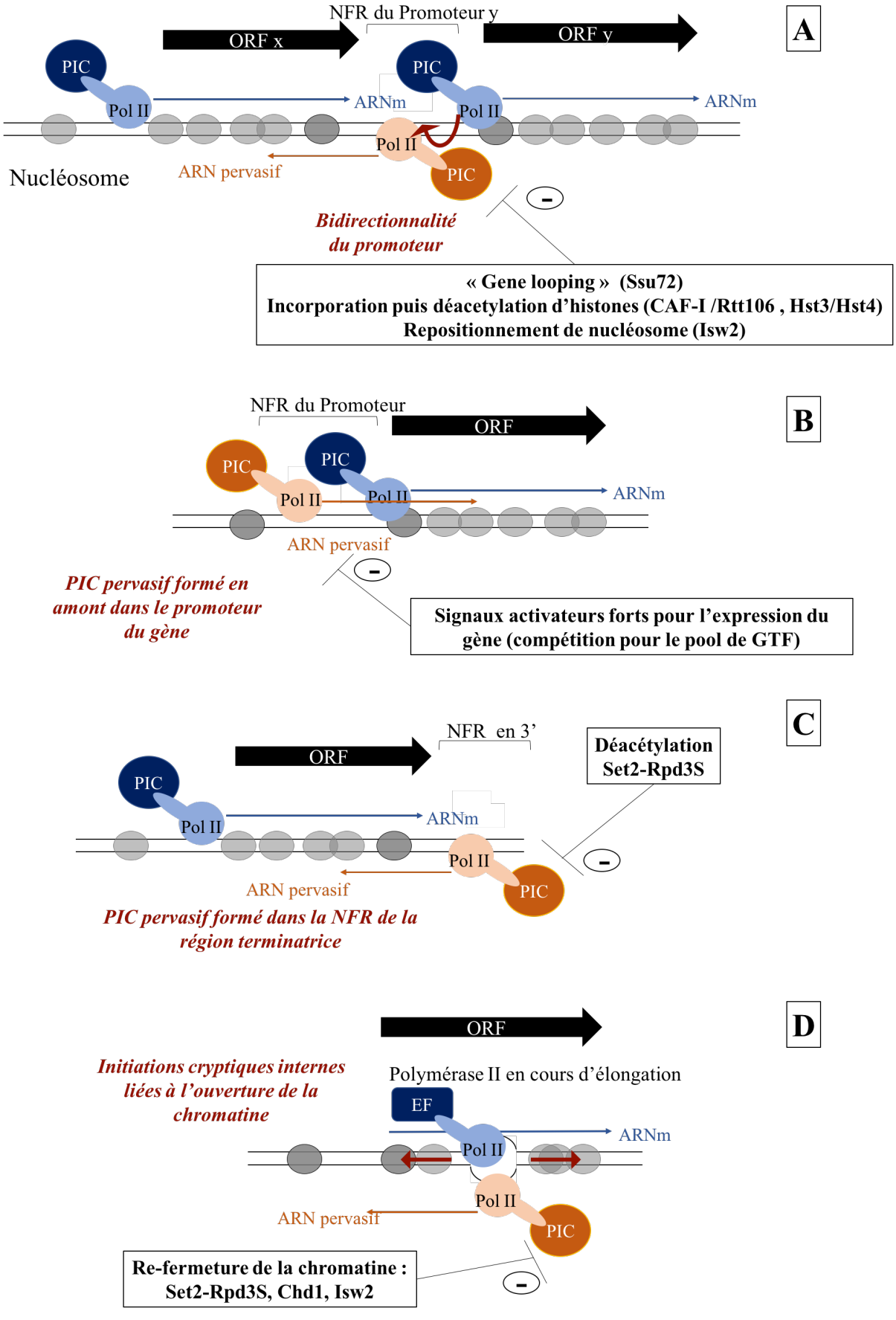


Figure 10 : Origines et contrôles de la transcription pervasive

Malgré la diversité des contrôles de la transcription pervasive au stade de l'initiation, leur efficacité reste partielle, et la cellule dispose d'un autre contrôle transcriptionnel pour arrêter précocement la transcription pervasive qui a pu être initiée et éviter qu'elle n'interfère avec la transcription génique.

2. Contrôle nucléaire conduisant à la dégradation des transcrits cryptiques

Le système de contrôle de la transcription pervasive en aval de l'initiation transcriptionnelle fait intervenir la voie de terminaison Nrd1-Nab3-Sen1, ou NNS, qui est également utilisée pour la terminaison des snoRNA.

Cette terminaison est permise par les facteurs Nrd1 et Nab3 qui reconnaissent spécifiquement des petites séquences sur l'ARN transcrit (GUAA/G pour Nrd1 et UCUUG pour Nab3). L'état de phosphorylation du CTD est déterminant dans le recrutement du complexe NNS. En effet, Nrd1 interagit directement avec le CTD S5P *via* son domaine CID. Le gradient de phosphorylation du CTD impacte donc fortement le choix de la voie utilisée pour la terminaison de la transcription, avec une marque S5P qui favorisera plutôt la terminaison NNS, tandis que la marque S2P favorisera la terminaison CPF-CF (Gudipati *et al*, 2008).

Nrd1 et Nab3 forment un complexe avec l'hélicase ADN Sen1 qui, en interaction avec l'ARN natif et grâce à une hydrolyse d'ATP, conduit à une dissociation du complexe de transcription de la matrice ADN. Le facteur Nrd1 échangera ensuite son interaction avec le CTD pour interagir avec le facteur Trf4 appartenant au complexe Trf4-Air2-Mtr4 (TRAMP) (Tudek *et al*, 2014). La polyadénylation du transcrit par le complexe TRAMP permet d'en faire une cible pour la dégradation par l'exosome nucléaire (Figure 9) (Porrua *et al*, 2016, 2012; Colin *et al*, 2011).

L'exosome est un complexe protéique permettant la maturation et la dégradation des ARN de l'extrémité 3' à l'extrémité 5'. Le cœur de l'exosome est composé d'une structure en anneau ou « PH ring » formée par les sous-unités Rrp41, Rrp42, Rrp43, Rrp45, Rrp46 et Mtr3. L'anneau est stabilisé par une coiffe protéique ou « S1-KH cap » contenant les sous-unités Rrp4, Csl4 et Rrp40 qui participe à la fixation à l'ARN et la reconnaissance des substrats. Cette coiffe permet à l'ARN de passer à travers le tunnel formé par l'anneau, en se positionnant à son entrée. En plus de ces sous-unités structurales, l'exosome contient des sous-unités catalytiques représentées par les exonucléases. Dis3 (ou Rrp44) est commune,

chez la levure bourgeonnante, aux exosomes nucléaire et cytoplasmique, alors que Rrp6 qui est spécifique à l'exosome nucléaire (Revues par Sloan *et al*, 2012; Kilchert *et al*, 2016) (Figure 10).

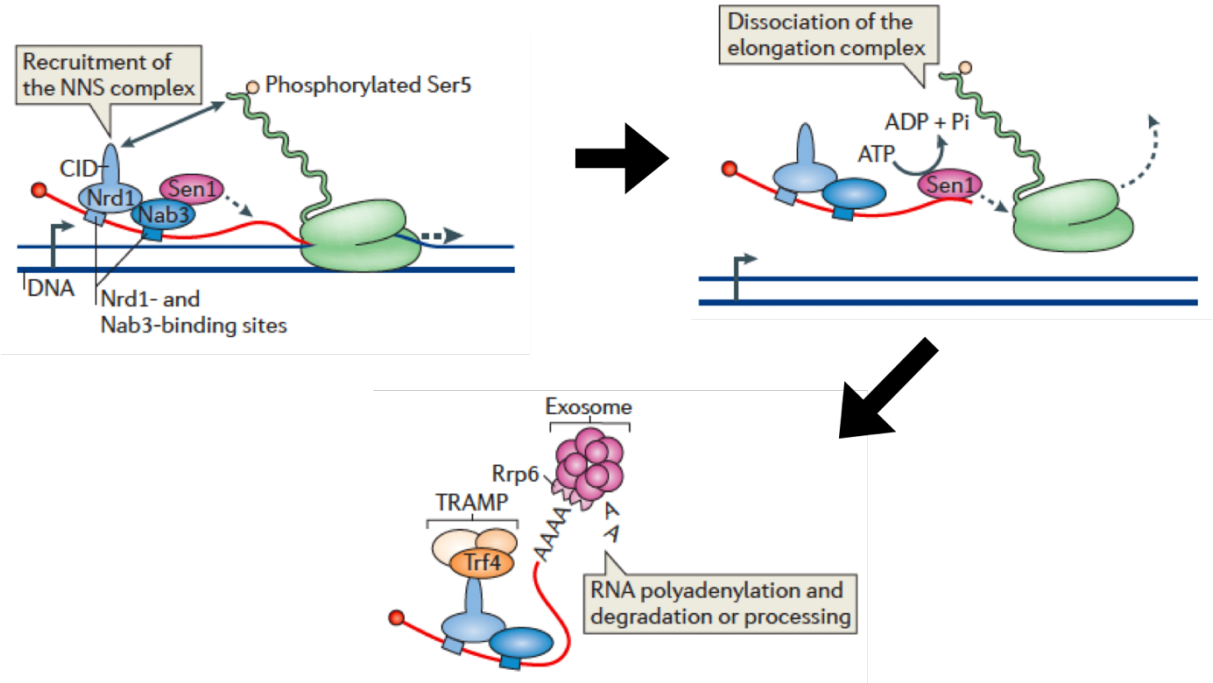


Figure 11 : Voie de terminaison Nrd1-Nab3-Sen1 (Adaptée de Porrua & Libri, 2015)

La dégradation par l'exosome nucléaire se limite aux ARN peu structurés. Le complexe TRAMP est donc important pour l'ajout de polyA non structurés par Trf4. Cette queue oligo-adénylée constitue une prise pour l'Hélicase Mtr4 qui peut ensuite déplier les structures ARN du transcrit ciblé et le rendre assimilable pour l'exosome nucléaire (Kilchert *et al*, 2016).

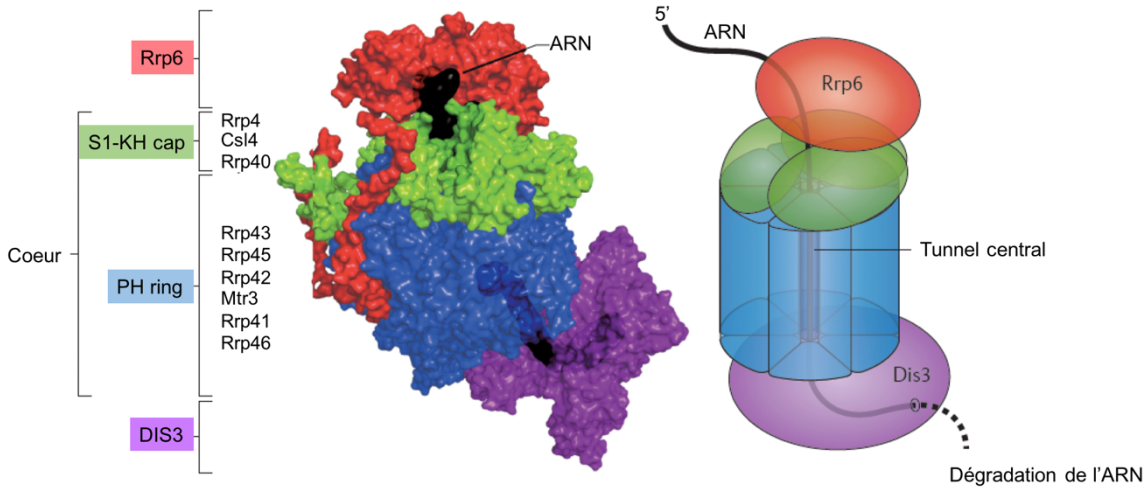


Figure 12 : Structure de l'exosome nucléaire (Adaptée de Kilchert et al, 2016)

La voie de terminaison NNS constitue un système de surveillance crucial pour restreindre la transcription pervasive déjà initiée et éviter ainsi des phénomènes d'interférence avec l'expression des gènes. En effet, il a été montré que le complexe Nrd1-Nab3-Sen1 a d'avantage tendance à se fixer sur les ARN pervasifs, car les ARNm sont généralement appauvris en motifs NNS (Jensen *et al*, 2013; Schulz *et al*, 2013; Porrua & Libri, 2015). La présence de « clusters » de sites de reconnaissance par Nrd1-Nab3 dans la région 5' des ARN non-codants (ncARN) est en partie responsable de l'hétérogénéité d'extrémité 3' générée par cette voie de terminaison (Wyers *et al*, 2005).

Les ncARN peuvent cependant échapper à la surveillance nucléaire de la transcription pervasive. Dans ce cas, les transcrits sont généralement plus allongés et terminés par la voie CPF-CF. Une étude a d'ailleurs montré que la plupart des transcrits pervasifs allongés utilisent les sites de terminaisons de transcription génique les plus proches (Uwimana *et al*, 2017). Les longs ARN non-codants (lncARN) générés sont ensuite exportés vers le cytoplasme.

De même que tous les ncARN ne sont pas tous terminés par la voie NNS, les gènes peuvent utiliser un mode de terminaison alternatif à la voie CPF-CF pour réguler le niveau d'ARNm disponible dans la cellule. Au cours de la transcription du gène *NRD1*, le choix de la voie de terminaison est déterminé par le niveau en protéine Nrd1. Si ce niveau est élevé, la terminaison précoce NNS sera favorisée et conduira à la dégradation du transcrit. A des niveaux faibles de protéine Nrd1, la terminaison CPF-CF sera favorisée, et le gène pourra être exprimé (Arigo *et al*, 2006; Porrua & Libri, 2015). De la même façon, le niveau d'expression de gènes tels que *URA2*, *FKS2* et *RPL9B* est régulé en fonction des besoins de la cellule par le choix de terminaison (Thiebaut *et al*, 2008; Kim & Levin, 2011; Gudipati *et al*, 2012).

3. Surveillance cytoplasmique des ARN : la voie du Nonsense-mediated mRNA decay (NMD)

Les lncARN qui ont échappé au contrôle nucléaire de la transcription pervasive sont exportés vers le cytoplasme. Ces ARN pervasifs peuvent y être délétères : s'ils s'accumulent et entrent dans un cycle traductionnel, ils sont susceptibles de mobiliser les ribosomes ou leurs produits de traduction peuvent être toxiques, par exemple s'ils saturent les machineries de surveillance des protéines mal repliées. Cependant, le fait qu'ils ne possèdent pas de longues phases codantes mais seulement de très courtes phases ouvertes aléatoires en font des cibles

privilégiées pour le contrôle-qualité cytoplasmique constitué par la voie du « Nonsense-mediated mRNA Decay » (NMD).

La voie du NMD intervient après l'export des ARN transcrits vers le cytoplasme et est couplée à la traduction par les ribosomes. On peut distinguer au moins 4 étapes successives pour la traduction : l'initiation, l'élongation, la terminaison et le recyclage des ribosomes. Pendant l'initiation de la traduction, la petite sous-unité ribosomale (40S) se lie à l'ARNt spécifique à la méthionine et à l'ARNm. Après sélection du triplet AUG ou codon START d'initiation de la traduction sur l'ARNm, cette petite sous-unité est rejointe par la grosse sous-unité ribosomale (60S) pour former le ribosome fonctionnel (80S). Pendant la phase d'élongation, le ribosome utilise les ARNt pour lire les codons de l'ARNm et ajouter les acide-aminés successifs à la chaîne peptidique naissante. La terminaison de la traduction comprend la reconnaissance d'un codon STOP sur la séquence de l'ARNm, la libération du peptide synthétisé et la dissociation du ribosome pour son recyclage (Dever *et al*, 2016).
L'absence d'une capacité codante suffisante de l'ARN traduit, une rétention d'intron, ou encore la présence d'une uORF, notamment, peuvent conduire à la reconnaissance d'un codon STOP précoce PTC pour « Premature Termination Codon » par le ribosome, et à un défaut de terminaison traductionnelle.

Chez la levure, le modèle du faux 3' UTR a été proposé pour expliquer comment les codons PTC sont distingués des codons STOP normaux. Dans ce modèle, la terminaison normale de la traduction fait intervenir une interaction entre la protéine liée à la queue poly-A, (PAPB, Pab1 chez la levure), et le facteur eRF3 lié au ribosome en cours de terminaison. Or, lorsqu'un PTC est présent, le ribosome en cours de terminaison est éloigné de la PAPB dont la concentration locale est donc plus faible que pour un ARNm normal. Upf1, une hélicase constituant l'un des facteurs centraux du NMD, peut alors, en association avec ses régulateurs Upf2 et Upf3, entrer en compétition favorable avec PAPB pour interagir avec eRF3 (Amrani *et al*, 2004; Brogna & Wen, 2009). Par ailleurs, la région en amont du codon STOP semble aussi jouer un rôle dans la sélection de cibles du NMD puisque, pour un 3' UTR de taille donnée, des transcrits avec une courte phase ouverte de lecture seront préférentiellement reconnus par Upf1 (Decourty *et al*, 2014).

Les transcrits ciblés par la voie du NMD sont ensuite décoiffés par les facteurs Dcp1-Dcp2 puis dégradés par l'exonucléase cytoplasmique 5'-3' Xrn1. Une voie alternative, impliquant

une déadénylation puis une dégradation de 3'en 5 par l'exosome cytoplasmique activé par le complexe Ski, a également été décrite (He & Jacobson, 2015; Mitchell & Tollervey, 2003). Le peptide naissant, potentiellement toxique pour la cellule, sera par ailleurs dégradé par le protéasome (Kuroha *et al*, 2009).

La voie du NMD représente un contrôle-qualité général pour débarrasser la cellule des transcrits incorrects. Elle constitue de fait un mécanisme cytoplasmique efficace de contrôle de la transcription pervasive, éliminant les transcrits pauvres en capacité codante et ayant échappé aux contrôles transcriptionnels dans le noyau (Malabat *et al*, 2015).

4. Mise en évidence des transcrits pervasifs dans la cellule

Les transcrits pervasifs ont été définis chez *Saccharomyces cerevisiae* selon les contrôles qu'il a fallu invalider pour leur détection. En 2003, une étude a pour la première fois mis en évidence une transcription pervasive décrite comme le résultat de phénomènes d'initiations transcriptionnelles aberrantes à partir de régions codantes, dans des mutants *spt6* (Kaplan *et al*, 2003).

Ce n'est qu'en 2005 qu'une première classe de 1486 transcrits cryptiques appelés CUT pour « Cryptic Unstable Transcripts », fut révélée en affectant la dégradation par l'exosome nucléaire par l'utilisation d'un mutant *rrp6* (Wyers *et al*, 2005). Une autre classe d'ARN pervasifs appelés SUT pour « Stable Unannotated Transcripts », fut opposée en 2009 à celle des CUT car leur stabilité est suffisante pour permettre leur détection dans une souche sauvage (Xu *et al*, 2009). Cette première description de 847 SUT avait déjà montré l'existence d'un certain chevauchement entre les CUT et les SUT (12% des SUT), confirmé par la suite dans une étude montrant la sensibilité de certains SUT à la dégradation nucléaire (Marquardt *et al*, 2011). Par ailleurs, il avait été découvert que les SUT et principalement les CUT pouvaient être à l'origine de formes étendues d'ARN pervasifs formées en l'absence de Nrd1, appelées respectivement eSUT et eCUT.

En 2011, le laboratoire d'Antonin Morillon a pu mettre en évidence 1658 XUT pour « Xrn1-sensitive Unstable Transcripts » qui sont détectés en l'absence de l'exonucléase cytoplasmique Xrn1, effecteur de la voie de surveillance cytoplasmique du NMD (van Dijk *et al*, 2011). La sensibilité de la majorité des SUT à Xrn1 a par ailleurs été montrée. Le mutant *upf1*Δ est affecté de la même manière dans la voie de surveillance cytoplasmique des ARN

pervasifs et permet également une révélation des XUT (Malabat *et al*, 2015; Wery *et al*, 2016).

L'extension transcriptionnelle pervasive dépendante de Nrd1 fut décrite de façon exhaustive en 2013 par le laboratoire de Patrick Cramer. La déplétion nucléaire de Nrd1 conduit en effet à la mise en évidence de 1526 NUT pour « Nrd1 Unterminated Transcripts », incluant des extensions en 3' de 44% des CUT, 37% des SUT et 620 XUT (Schulz *et al*, 2013).

L'étude du recouvrement entre ces classes de transcrits, des mécanismes à l'origine de la transcription pervasive ainsi que de son contrôle dans la cellule a permis de prendre du recul par rapport à la définition des ARN non-codants. En résumé, une initiation pervasive peut être ciblée dans un premier temps par une terminaison précoce couplée à une dégradation nucléaire. Cependant une fraction de cette transcription échappant au contrôle de la voie NNS, est allongée et terminée par la voie CPF-CF. Le lncRNA mature exporté ensuite vers le cytoplasme peut s'associer aux ribosomes car il possède des codons ATG et STOP aléatoires formant de petites ORF, mais son absence de capacité codante substantielle conduira à la reconnaissance d'un PTC par la machinerie du NMD et en fera une cible pour cette voie de dégradation cytoplasmique (Figure 11). Cette coopération entre les voies de dégradation nucléaire et cytoplasmique permet ainsi une élimination efficace des transcrits pervasifs, et explique les effets additifs des mutations de *RRP6* et *UPF1* sur le niveau de nombreux transcrits pervasifs (Malabat *et al*, 2015; Marquardt *et al*, 2011).

Dans le cadre de l'étude de la fonctionnalité des transcrits pervasifs au niveau transcriptionnel, on fera une distinction entre les transcrits pervasifs majoritairement ciblés par un contrôle nucléaire et les lncRNA (quelle que soit leur dénomination). En effet, les transcrits principalement ciblés dans le noyau par une terminaison précoce seront plus courts, et auront un très faible potentiel régulateur pour l'expression des gènes. Les transcrits qui échapperont majoritairement à ce contrôle nucléaire sont potentiellement plus régulateurs, car leur transcription implique des séquences plus étendues et a donc potentiellement plus d'impact sur la structure de la chromatine.

L'utilisation du double mutant *rrp6*Δ*upf1*Δ permet de révéler la quasi-totalité des ARN transcrits naturellement par les cellules (Malabat *et al*, 2015). Ce mutant est donc intéressant dans la mesure où l'on voudrait mesurer ces transcrits pour étudier leur régulation.

Cependant, pour l'étude de leur rôle au niveau transcriptionnel, le mutant *rrp6*Δ*upf1*Δ présente un problème potentiel. En effet, il a été montré que la délétion de *RRP6* peut avoir en elle-même un effet négatif sur la terminaison de transcription NNS dépendante (Fox *et al*, 2015; Castelnuovo *et al*, 2014), possiblement parce que l'accumulation des cibles du NNS pourrait saturer cette machinerie (Tommaso Villa et Domenico Libri, communication personnelle). Les conséquences de défauts de terminaison dans le mutant *rrp6*Δ ne pouvant pas être distinguées des effets potentiels de la transcription pervasive naturelle, des expériences de séquençage d'ARN en cours de transcription par la polymérase par la méthode « Net-Seq » (Churchman & Weissman, 2011) seront plus indiquées pour l'étude de la fonctionnalité de la transcription non-codante. En effet, elles permettent de mesurer les transcrits avant une dégradation potentielle sans toutefois affecter le mécanisme général de transcription. Alternativement, on pourra étudier la fonctionnalité de la transcription pervasive dans un simple mutant *upf1*Δ, suffisant pour révéler un très grand nombre des lncRNA.

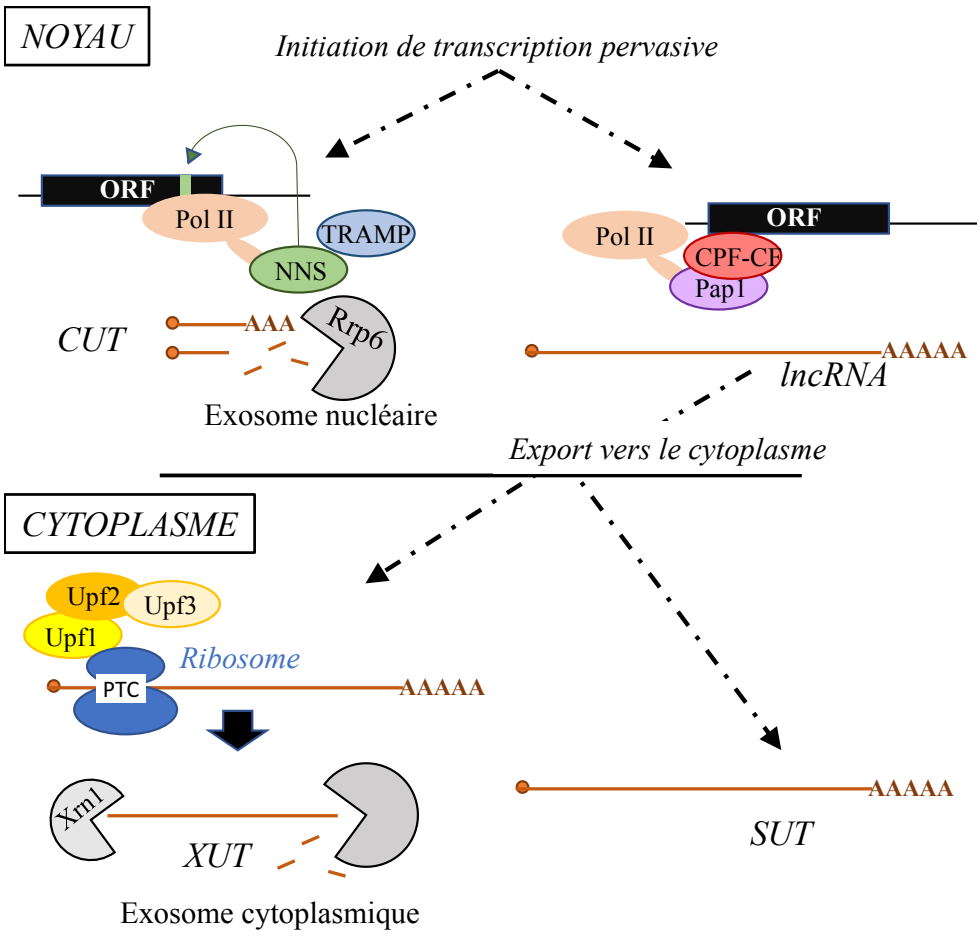


Figure 13 : Devenir cellulaire de la transcription pervasive

III. Rôles de la transcription pervasive

Diverses études de la transcription pervasive ont contribué à la compréhension de sa biogenèse, et de son devenir dans la cellule. Cependant, la question de sa fonctionnalité globale subsiste. On distinguera deux différents types de fonctionnalité potentielle de la transcription pervasive : (*i*) en *trans*, c'est à dire que c'est la molécule d'ARN non-codant qui joue un rôle potentiellement fonctionnel, rôle qui sera susceptible d'être réprimé par la dégradation de l'ARN (*ii*) en *cis*, c'est la transcription pervasive elle-même qui aura un impact, que l'ARN produit soit dégradé ou non en aval.

1. Fonctionnalité en *trans*

La recherche de système similaire à celui du RNAi chez *Saccharomyces cerevisiae* a suggéré le rôle régulateur en *trans* d'un ARN non-codant. En effet, la délétion de *XRN1*, en permettant une forte stabilisation du CUT/XUT antisens RTL, conduit à une répression de l'expression du transposon TY1. Cette répression est induite par le recrutement ARN-dépendant de l'HMT Set1, dont les marques épigénétiques sont reconnues par une HDAC à l'origine de la répression transcriptionelle (Berretta *et al*, 2008). Dans la continuité de cette étude, une analyse de l'occupation par la polymérase II dans un mutant *xrn1*Δ a pu montrer que la proportion de gènes négativement régulés en l'absence de xrn1 est plus importante pour ceux qui sont associés à un antisens. Il est notamment montré que le gène *TIR1* est régulé par un XUT antisens, l'interruption de cet antisens conduisant à une augmentation de l'ARNm associé. La répression du gène par le transcrit non-codant ferait intervenir l'HMT Set1 (van Dijk *et al*, 2011). Cependant, il n'y a pas de preuve directe que cet antisens exerce uniquement une action en *trans*, et que sa transcription elle-même n'est pas régulatrice.

Le même laboratoire a par ailleurs mis en évidence le rôle potentiel des XUT dans la régulation de l'expression génique au niveau de la traduction. Cette étude a montré qu'environ 80% des XUT antisens, une fois stabilisés par la délétion de *XRN1*, peuvent former des structures ARN double-brin en s'hybridant aux ARNm associés. Cette hybridation serait négativement régulatrice pour la traduction car elle empêcherait la fixation des ribosomes, tant sur l'antisens que sur l'ARNm (Wery *et al*, 2016).

Malgré ces deux exemples, il semblerait de la levure bourgeonnante *Saccharomyces cerevisiae* se distingue des eucaryotes complexes par la très faible importance des régulations

en *trans* permises par des ARN pervasifs. En effet, les cas décrits de régulations liées à la transcription pervasive montrent majoritairement un impact de l'acte transcriptionnel en lui-même plutôt que de la molécule d'ARN produite.

2. Fonctionnalité en *cis*

La transcription pervasive joue en *cis* un rôle global, lié à des modifications co-transcriptionnelles de la chromatine. En particulier, une étude globale a mis en évidence la signature particulière de la transcription pervasive antisens sur la structure de la chromatine, bien distincte de celle de la transcription génique. Pour un même niveau d'expression de l'ARNm (expression « sens »), la transcription en antisens est en effet à l'origine d'une présence accrue en facteurs remodeleurs de la chromatine et en chaperonnes d'histone pour augmenter la densité en nucléosomes, et réduire l'étendue des NFR dans les régions promotrices des gènes. Sur la région codante, la transcription antisens accentue le renouvellement des histones (Figure 12). Ces modifications structurales de la chromatine augmentent ainsi le nombre de conformations possibles, et favoriseraient ainsi le « bruit » transcriptionnel, c'est à dire la variabilité d'expression entre des cellules individuelles (Murray *et al*, 2015).

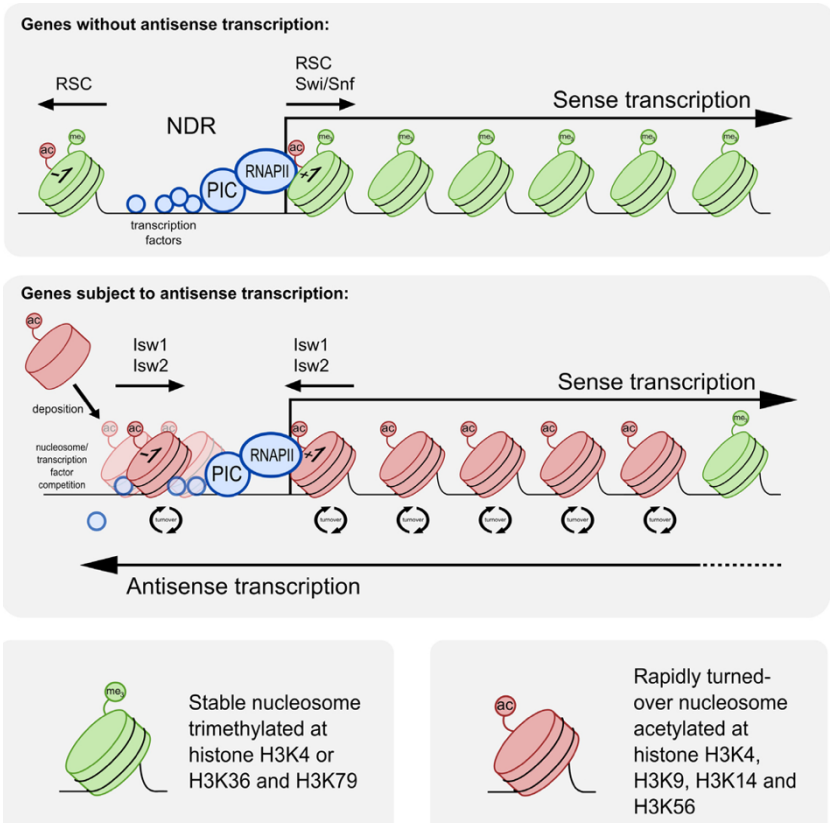


Figure 14 : Impact de la transcription antisens sur la chromatine dans les régions promotrices et codantes des gènes (Murray et al, 2015)

Des cas de régulation positive par la transcription pervasive ont été démontrés. Par exemple, le lncARN initié en 3' de PHO5 et recouvrant son promoteur a été impliqué dans la cinétique d'activation du gène, probablement par l'intermédiaire des modifications de la chromatine liées à la transcription du lncARN (Uhler *et al*, 2007). L'autre exemple concerne l'induction du gène *CDC28* lors de stress osmotiques. La transcription SAGA-dépendante de l'antisens de *CDC28* est spécifiquement induite par l'activateur transcriptionnel Hog1 lors de stress osmotiques. Un phénomène de « gene looping » (dépendant de Ssu72) permet alors le transfert de Hog1 au nucléosome +1 puis le recrutement du complexe RSC au niveau du promoteur génique. Le remodelage induit conduit ensuite à l'induction de l'expression de CDC28 (Nadal-Ribelles *et al*, 2014).

L'HDAC Set3 a été impliquée dans des régulations de gènes par les modifications de la chromatine liées à la transcription non-codante. Une étude a décrit ses effets divers, pouvant être positifs ou négatifs pour la transcription des ARNm. Par exemple, la transcription d'un ncRNA sens en amont du TSS du gène *DCI1* conduit à la déacétylation du promoteur génique par Set3, conduisant à l'inhibition de son expression ou au retard de son induction (Kim *et al*, 2012).
La transcription de lncARN en antisens de promoteurs de gènes, tels que *YNL195C* ou *GLO4*, permet également de recruter le complexe Set3C *via* les marques H3K4me2 déposées co-transcriptionnellement (par l'HMT Set1). La déacétylation par Set3C atténue la réponse transcriptionnelle à une variation environnementale (Kim *et al*, 2012).
Certaines transcriptions antisens initiées en 3' des gènes permettent, grâce à l'intervention de Set3C, d'atteindre un niveau d'induction maximal de l'expression des gènes pour l'adaptation à une variation environnementale. Dans ce cas, les auteurs supposent que Set3C agit pour réprimer une transcription non-codante qui serait inhibitrice pour l'expression des gènes (Kim *et al*, 2012). Par exemple, le gène *ADH1* codant pour une alcool déshydrogénase est réprimé lors d'une carence en zinc par la transcription pervasive de *ZRR1*, qui est initiée en amont et conduit à l'occlusion de son promoteur (Bird *et al*, 2006). L'observation d'une plus forte induction de ce gène dans un mutant *set3*Δ pourrait donc être liée à la répression de *ZRR1*.

Des facteurs de remodelage jouent un rôle dans le mécanisme d'interférence transcriptionnelle régulant certains gènes comme *SER3* qui est régulé par le lncARN *SRG1* initié en amont dans la même orientation. A des concentrations élevées en sérine, *SRG1* est

induit par l'activateur Cha4 reconnaissant une UAS de son promoteur. Cha4 est responsable du recrutement des remodeleurs de la chromatine SAGA et Swi/Snf qui sont des co-activateurs pour la transcription de *SRG1* (Martens *et al*, 2004, 2005). Un remodelage induit par les facteurs de remodelage Spt6 et Spt16 (au sein du complexe FACT) participe à l'ouverture de la chromatine pour la transcription non-codante mais aussi à son rétablissement après le passage de la polymérase. Ce rétablissement de la structure chromatinienne inclut le repositionnement des nucléosomes préalablement dissociés par Swi/Snf sur le promoteur du gène *SER3* (Hainer *et al*, 2011, 2012), conduisant ainsi à sa répression. Plus récemment, le rôle du complexe FACT a été montré, notamment pour empêcher l'initiation de transcrits nommés « Spt16-suppressed noncoding transcripts ». La délétion de SPT16 conduirait à une forte instabilité des nucléosomes après transcription par la polymérase II (Feng et al, 2016).

L'HDAC Rpd3S intervient dans le mécanisme d'interférence transcriptionnelle responsable de la régulation du locus *GAL1-GAL10*. La transcription du lncARN antisens au gène *GAL10* est induite par Reb1 et responsable de la répression du gène en l'absence de galactose, *via* le recrutement de la voie Set2-Rpd3S (Houseley *et al*, 2008). Ce même lncARN est dans l'orientation sens par rapport à *GAL1*, et participe à sa répression *via* un recrutement de Rpd3L, cette fois-ci par la marque H3K4me2 déposée par Set1 (Pinskaya *et al*, 2009). Une étude plus récente a démontré que, bien que le lncARN soit transcrit de façon constitutive, il n'a qu'un statut de bruit transcriptionnel sans fonction lorsque la croissance se fait en galactose. Sa transcription devient régulatrice en glucose, afin d'empêcher une fuite transcriptionnelle des gènes censés être éteints (Lenstra *et al*, 2015).

Le gène *FLO11* est régulé par deux transcrits non-codants antagonistes *ICR1* et *PWR1* qui sont l'un ou l'autre favorisés selon les conditions environnementales. *ICR1* est initié dans l'orientation sens, bien en amont de *FLO11*. Sa transcription, allant au-delà du promoteur de *FLO11*, conduit à son occlusion et empêche grâce à Rpd3L la fixation des activateurs Flo8 et Sfl1, conduisant ainsi à la répression génique. Le transcrit *PWR1* est lui initié en antisens d'*ICR1* et interfère avec son expression (Bumgarner *et al*, 2009, 2012).

Un autre locus, contenant deux gènes *HMS2* et *BAT2* organisés en tandem, est régulé par la transcription non-codante, bien que le mécanisme de régulation n'ait pas encore été élucidé. La transcription du gène *HMS2* interfère avec celle de *BAT2* et la transcription du SUT non-codant antisens à *HMS2* limite son niveau d'expression et dé-réprime de ce fait l'expression

de *BAT2*. Le modèle décrit est encore plus complexe puisque *HSM2* peut également utiliser le terminateur de BAT2 pour former un transcrit bi-cistronique (Nguyen *et al*, 2014).

Le déclenchement du processus de sporulation est contrôlé par deux gènes, *IME1* et *IME4*, tous deux régulés par un transcrit pervasif. En effet, le gène *IME4* est régulé négativement par l'antisens non-codant RME2, dont la répression par l'hétérodimère a-α permet l'induction d'*IME4* pour l'entrée en méiose. *IME1* est lui réprimé par l'action de Set2 et Set3, recrutés par la transcription du XUT sens IRT1 initié en amont (van Werven *et al*, 2012).

Un rôle de suppression de la mémoire transcriptionnelle a par ailleurs été démontré pour un transcrit pervasif associé au gène de l'endonucléase HO. En effet, une exposition des cellules au facteur α déclenche une cascade d'activation (MAP kinase) aboutissant à la phosphorylation du facteur de transcription Ste12, activateur de la transcription de pHO-lncARN. Sa transcription en sens, initiée en amont de HO permet de retirer les facteurs SBF fixés sur son promoteur et supprimer ainsi la mémoire d'une activation transcriptionnelle antérieure, après un arrêt du cycle cellulaire. La transcription non-codante permet donc ainsi d'inhiber l'expression de HO lorsque les cellules reprennent leur division (Yu *et al*, 2016).

Une étude du laboratoire de Lars Steinmetz décrit la fonction d'antisens issus de la bidirectionnalité des promoteurs dans des régions contenant des gènes en tandem. En effet, la transcription non-codante pourrait y agir pour étendre des signaux régulateurs à des gènes voisins qui doivent être co-régulés par des variations environnementales. Par exemple, la répression en *cis* du gène *SUR7* par *SUT719*, en antisens, a lieu lorsque le transcrit non-codant divergeant du promoteur de *GAL80* est induit pendant une croissance en galactose. Par ailleurs, ce laboratoire a pour la première fois formulé l'hypothèse d'une co-régulation dépendante d'un certain seuil d'expression, entre la transcription antisens et la transcription codante (Xu *et al*, 2011). Ce type de régulation permettrait une plus grande variabilité d'expression pour les gènes associés à des antisens.

La régulation du gène *PHO84* par la transcription antisens a été étudiée très en détail. En l'absence de Rrp6, la fréquence d'extension de la transcription antisens à *PHO84* augmente en raison d'une diminution de la stabilité d'association du complexe NNS ciblant cette transcription non-codante. Cela conduit à l'allongement de la transcription antisens jusqu'au niveau du promoteur génique, permettant alors le recrutement de l'HDAC Hda1. L'HMT Set1

favorise l'élongation de la transcription antisens, mais aussi le recrutement d'Hda1 par celle-ci. La déacétylation dont Hda1 est à l'origine conduit à la répression de la transcription génique. Cette régulation par interférence transcriptionelle permet d'assurer une inhibition de l'expression du gène lorsque les signaux activateurs (par Pho4) ne sont pas suffisants, afin d'éviter des phénomènes de « fuite transcriptionelle » lorsque le gène est normalement éteint (Camblong *et al*, 2007, 2009; Castelnuovo *et al*, 2013). Une étude plus globale réalisée par ce laboratoire visait à analyser le rôle de la terminaison NNS et des modifications de la chromatine dans la répression génique. Dans cette étude, l'équipe de Francoise Stutz a pu montrer que la sensibilité des gènes à l'interférence transcriptionnelle était en partie conditionnée par la capacité de leur transcription antisens à échapper à la terminaison NNS, et que différentes voies de modifications de la chromatine sont impliquées dans le mécanisme de répression (Castelnuovo *et al*, 2014).

Tableau 1 : Récapitulatif des cas décrits de régulation des gènes par la transcription pervasive

Gène régulé	**ARN pervasif sens/antisens**	**lncARN activement régulé**	**Régulation cis/trans**	**Régulation pour l'expression génique**	**Facteurs impliqués dans la régulation**	**Références**
TY1	Antisens (RTL)	Non	*trans*	Négative	Set1, HDAC	Berretta *et al*, 2008
TIR1	Antisens	N.E.	*trans*	Négative	Set1	van Dijk *et al*, 2011
PHO5	Antisens	N.E.	*cis*	Positive		Uhler *et al*, 2007
CDC28	Antisens	Oui (induction par Hog1 - Stress osmotique)	*cis*	Positive	Ssu72	Nadal-Ribelles *et al*, 2014
DCI1	Sens	N.E.	*cis*	Négative	Set3	Kim *et al*, 2012
ADH1	Sens (ZRR1)	Oui (répression par Set3)	*cis*	Négative	?	Kim et al, 2012 Bird *et al*, 2006
YNL195C	Antisens	N.E.	*cis*	Négative	Set3	
GLO4	Antisens	N.E.	*cis*	Négative	Set3	Kim *et al*, 2012
SER3	Sens (SRG1)	Oui (induction par Cha4, SAGA, SWI/SNF - disponibilité en nucléotide)	*cis*	Négative	Spt6 , FACT (Spt16)	Martens *et al*, 2004, 2005

GAL1 /GAL10	Sens /Antisens	Non (transcrit constitutif)	*cis*	Négative	Set2-Rpd3S /Set1-Rpd3L	Houseley *et al*, 2008 Pinskaya et al, 2009 Lenstra et al, 2015
FLO11	Sens (ICR1)	N.E.	*cis*	Négative	Rpd3L	Bumgarner *et al*, 2009, 2012
HMS2	Antisens	N.E.	*cis*	Négative	?	Nguyen *et al*, 2014
IME1	Sens	Oui (induction par le facteur de transcription Rme1)	*cis*	Négative	Set2 Set3	van Werven *et al*, 2012
IME4	Antisens (RME2)	Oui (répression par l'hétérodimère a-$\propto$)	*cis*	Négative	?	Hongay *et al*, 2006 Gelfand *et al*, 2011
HO	Sens (pHO-lncRNA)	Oui (induction par le facteur de transcription Ste12	*cis*	Négative	?	Yu *et al*, 2016
SUR7	Antisens	Non (induction liée à l'induction de GAL80)	*cis*	Négative	?	Xu *et al*, 2011
PHO84	Antisens	N.E.	*cis*	Négative	Hda1 Set1	Camblong *et al*, 2007, 2009 Castelnuovo *et al*, 2013 Castelnuovo et al, 2014

La construction d'un système rapporteur pour étudier l'effet de la transcription antisens sur le niveau des protéines a pu montrer que la transcription d'un quart des SUT antisens peut affecter l'expression du gène associé, de façon positive ou négative. Les régulations sont généralement négatives et de faible amplitude. Les gènes les plus faiblement exprimés sont relativement plus sensibles à la transcription antisens, d'autant plus que celle-ci va au-delà du TSS du gène associé (Huber *et al*, 2016).

Chez les eucaryotes complexes, des régulations en *cis* par les antisens non-codants ont également été décrites. Par exemple, le transcrit aberrant *LUC7L* conduit chez l'Homme à une méthylation des îlots CpG présents au niveau du promoteur du gène *HBA1* codant pour la chaîne d'hémoglobine α1. La transcription non-codante conduit donc à l'inhibition

transcriptionnelle du gène *HBA1*, responsable de maladies thalassémiques (Pelechano & Steinmetz, 2013). Cependant, les lncARN sont principalement fonctionnels en *trans* par leur capacité à recruter des complexes de modifications de la chromatine. Un exemple, connu de longue date, de régulation en *trans* par des transcrits non-codants est celui de l'inactivation du chromosome X chez les mammifères, impliquant notamment les ARN non-codants *XIST* et *TSIX*. Au début du développement embryonnaire, les deux chromosomes contiennent à la fois les transcrits *TSIX* à un fort niveau d'expression, et *XIST* à un niveau d'expression basal. La transcription de *TSIX* génère des modifications de la chromatine au niveau du promoteur de *XIST*, conduisant à sa répression. Au cours de la différenciation, la transcription de *TSIX* persiste sur un chromosome, et sa perte sur l'autre chromosome permet en revanche l'activation transcriptionnelle de *XIST* et sa diffusion. L'ARN *XIST* recrute ensuite des complexes de remodelage de la chromatine qui conduisent à l'inactivation du chromosome X (Revue par Lee, 2010).

En conclusion, chez *Saccharomyces cerevisiae*, de nombreuses études ont montré que la transcription pervasive avait un potentiel régulateur (Castelnuovo & Stutz, 2015). Dans la majorité des cas, contrairement aux eucaryotes complexes, l'ARN non-codant en lui-même n'est pas fonctionnel mais c'est la transcription de cet ARN qui joue un rôle régulateur. L'ARN non-codant est d'ailleurs ciblé par des mécanismes de dégradation et peut donc être considéré comme un sous-produit de ce mécanisme. La transcription pervasive antisens semble majoritairement avoir un effet répresseur pour l'expression génique, bien que les exemples analysés soient relativement peu nombreux, et impliqués dans des processus biologiques différents (Tableau 1) (sauf pour les régulations *d'IME1* et *IME4* par des transcrits pervasifs différents afin de contrôler la sporulation). La question de la fonctionnalité globale de cette transcription non-codante pour l'adaptation à une même variation environnementale reste donc ouverte. De plus, la coexistence de diverses voies de modifications de la chromatine pour la régulation des gènes par interférence transcriptionnelle (Tableau 1) appelle à des études supplémentaires de ce phénomène pour rechercher l'ensemble des facteurs potentiellement impliqués dans les mécanismes de répression, et identifier les différentes voies possibles.

IV. Projet de thèse

C'est autour de cette problématique liée à la transcription pervasive, à sa fonctionnalité et aux mécanismes de régulation des gènes qu'a été élaboré mon projet de thèse. Il s'est organisé autour de deux axes principaux :

- **L'étude des mécanismes d'interférence transcriptionelle chez *Saccharomyces cerevisiae*.**
- **L'étude de l'implication du phénomène d'interférence transcriptionelle, lié à la transcription antisens, dans la co-régulation de différents gènes.**

Durant la première partie de ma thèse, mon projet a été construit en continuité avec celui de mon Master, pendant lequel j'ai utilisé un système artificiel pour l'induction de multiples phénomènes d'interférence transcriptionnelle. Cet outil m'avait permis de rechercher l'implication de facteurs « candidats » dans le mécanisme de répression. Mon projet de thèse a dans un premier temps consisté à chercher ces facteurs par une approche moins biaisée, en développant un système rapporteur pour reproduire le phénomène d'interférence transcriptionnelle et réaliser des cribles génétiques. Ce système rapporteur a également été construit dans le but de l'associer avec différentes combinaisons de promoteurs géniques et cryptiques afin d'identifier des critères de sensibilité à l'interférence transcriptionnelle et de spécificité de facteurs impliqués dans la répression.

Dans la deuxième partie de ma thèse, j'ai réalisé des études transcriptomiques à grande échelle dans le but de comparer la transcription génique et la transcription non-codante. J'ai ensuite identifié une population de gènes vraisemblablement co-régulés par interférence transcriptionelle après une variation environnementale, et j'ai validé ce phénomène par des expériences d'interruption directionnelle de la transcription, ciblant à la fois la transcription non-codante antisens et celle des ARNm.

Résultats

I. Description des mécanismes d'interférence transcriptionelle chez *Saccharomyces cerevisiae*

1. Résultats précédents

Lors de mon stage de Master 2, j'ai pu utiliser un système d'induction de multiples phénomènes d'interférence transcriptionnelle à l'échelle du génome. En effet, l'utilisation du mutant tetNrd1, dans lequel l'expression de *NRD1* est réprimée par la doxycycline, permet l'induction simultanée de l'allongement de nombreux transcrits non-codants. Le même principe a été utilisé par la laboratoire de Patrick Cramer qui a révélé les NUT grâce à une séquestration transitoire de la protéine Nrd1 nucléaire dans le cytoplasme, par une technique appelée « Anchor-away » (Schulz *et al*, 2013). Au laboratoire, nous avons associé à la déplétion de Nrd1 une invalidation de la voie du NMD via une délétion *upf1*Δ, afin d'être en mesure de détecter les transcrits allongés qui seraient ensuite exportés vers le cytoplasme et ciblés par cette voie de contrôle-qualité des ARN. Toutes les souches de l'étude ont donc été cultivées pendant 11h dans un milieu additionné de doxycycline.

Des techniques de puces à ADN, bien que moins exhaustives et sensibles que celles de séquençage, ont été validées pour l'étude du phénomène d'interférence transcriptionnelle dans le mutant *tetNrd1nam7*Δ, et ont permis de définir deux classes de gènes particulières :

- 103 gènes dits « Cibles » qui sont associés à un transcrit antisens, allongé en l'absence de Nrd1.
- 76 gènes dits « Régulés » qui représentent une sous-catégorie de cette première classe. Leur antisens est induit au moins d'un facteur 2 et ils sont réprimés au moins d'un facteur 2 en l'absence de Nrd1. Ils sont donc très fortement réprimés par un phénomène d'interférence transcriptionnelle lié à l'allongement du transcrit antisens associé.

Dans le but de décrire les mécanismes d'interférence transcriptionnelle, j'ai testé l'implication dans la répression génique de différents facteurs candidats responsables de modifications de la chromatine : Set1, Hda1, Chd1 et Set3. Pour cela, j'ai délété ces facteurs dans le fond génétique *tetNrd1upf1*Δ*,* et analysé la conséquence de la délétion sur les transcriptions sens et

antisens. Plus particulièrement, j'ai étudié les gènes « Régulés » pour rechercher des cas où l'absence d'un facteur de modification de la chromatine ne permettait plus la répression du gène par l'allongement de l'antisens. Le facteur de modification Set2 avait déjà été testé par Frank Feuerbach dans le laboratoire, selon le même principe mais par des expériences de RNA-seq. Pour m'affranchir des effets de la délétion simple des facteurs de modification de la chromatine qui ne seraient pas liés à l'interférence transcriptionnelle, j'ai étudié l'effet de la mutation simple du facteur en parallèle de l'effet de cette mutation dans le fond *tetNrd1upf1*Δ, pour me focaliser sur des régulations spécifiques à l'interférence transcriptionnelle.

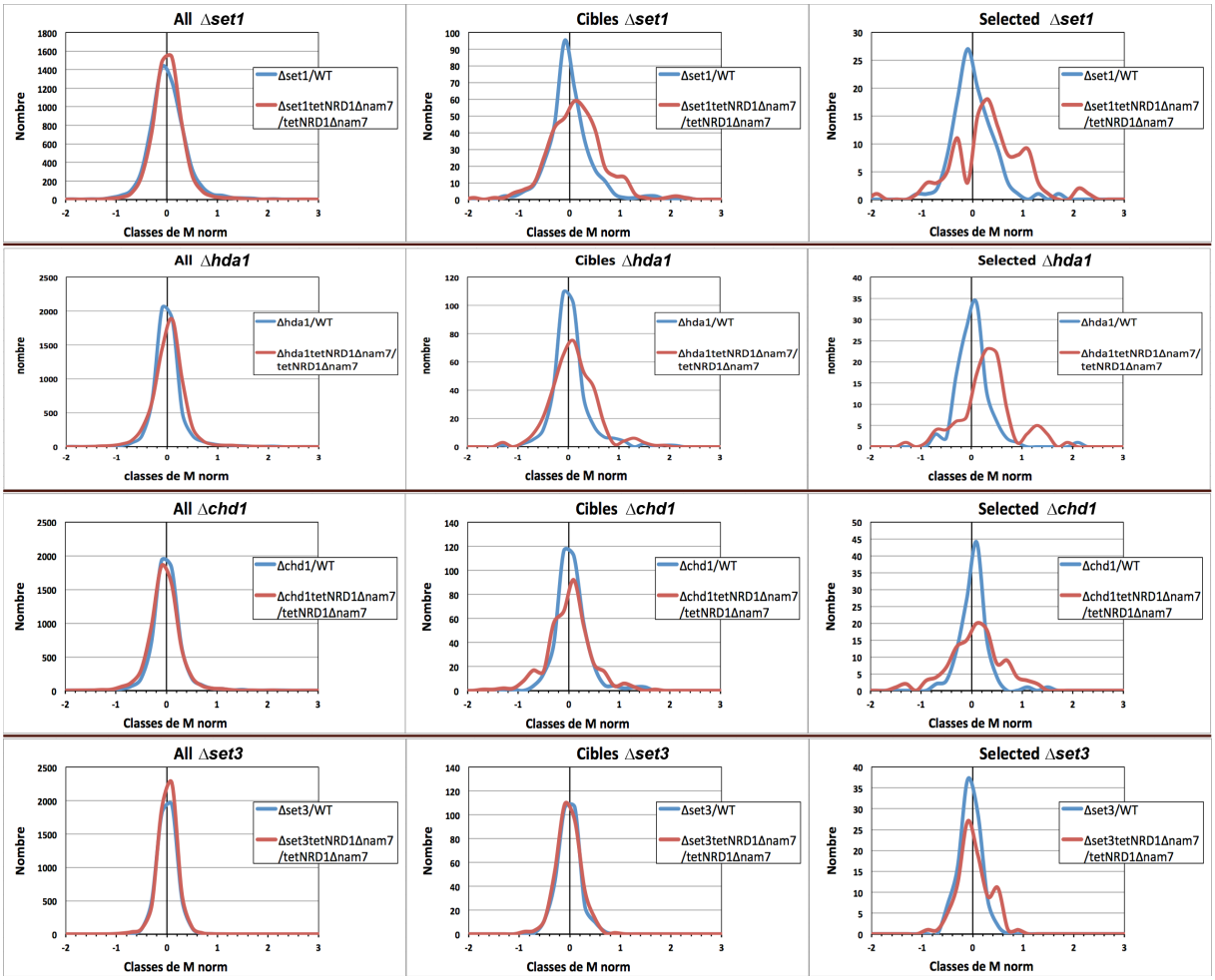


Figure 15: Distribution des gènes suivant l'effet de la délétion du facteur de modification de la chromatine sur le niveau d'ARNm (M norm = log2 (abondance Δfacteur */abondance)).*

L'effet global est montré pour chaque facteur testé,

pour tous les gènes (« All »), les gènes « Cibles », et les gènes « Régulés ».

*Les courbes en rouge représentent les effets de la délétion dans le fond tetNrd1upf1*Δ*, alors que les courbes en bleu représentent les effets de la délétion simple.*

*NAM7 est l'autre nom du gène UPF1 (*nam7Δ=upf1Δ*).*

Contrairement aux facteurs Chd1 et Set3, l'absence des facteurs Set1 et Hda1 semble avoir un effet, global mais faible, d'augmentation de l'expression des gènes associés à des antisens (Figure 15 comparaison entre « All » et « Cibles », effet spécifique du fond *tetNrd1*Δ*nam7*). Pour tous les facteurs testés, on peut détecter un effet plus fort de la délétion qui est spécifique aux gènes réprimés par interférence transcriptionnelle et limité à un petit nombre de gènes (Figure 15, comparaison entre « Cibles » et « Régulés »). Il y aurait donc deux niveaux de régulation des gènes par la transcription antisens impliquant des modifications de la chromatine. Un niveau de régulation global mais faible, et un niveau de régulation plus fort *via* un phénomène d'interférence transcriptionnelle utilisant une voie gène-spécifique.

Cet effet spécifique de la délétion des facteurs de modification de la chromatine a été analysé, en définissant un seuil d'abolition de la répression du gène par l'antisens, quand ce dernier est toujours produit. Pour cela, les gènes « Régulés » restaurés d'un facteur supérieur ou égal à 1,4 et pour lesquels l'antisens n'est pas diminué d'un facteur inférieur à 0,7 ont donc été classés et représentés selon le facteur de modification de la chromatine impliqué (Figure 16).

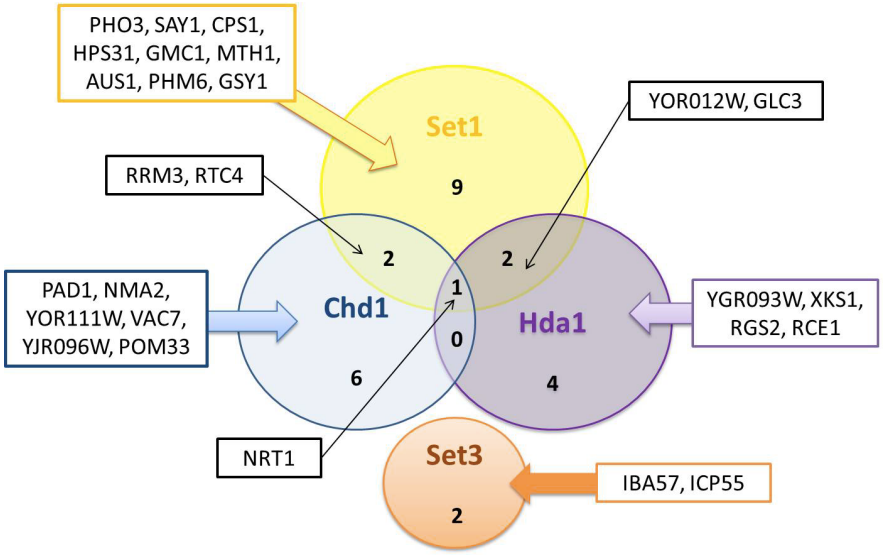


Figure 16 : Classement des gènes réprimés par interférence transcriptionnelle ("Régulés") selon le facteur de modification de la chromatine spécifiquement impliqué dans le mécanisme

Chd1, Set1, Hda1 et Set3 ont pu être impliqués dans le phénomène de répression pour 22 gènes « Régulés » sur les 76 (28%). Cela suggère qu'il pourrait y avoir d'autres voies de répression par interférence transcriptionnelle, impliquant d'autres facteurs de modification de la chromatine. De façon intéressante, on observe un recouvrement entre les groupes de gènes « Régulés » affectées par les délétions des différents facteurs testés, qui suggère une certaine redondance des voies impliquant ces facteurs.

Par ailleurs, cette étude utilisant l'outil du mutant *tetNdr1upf1*Δ a permis de mettre en évidence l'existence de gènes « insensibles » à l'interférence transcriptionnelle, pour lesquels l'allongement d'un antisens au-delà du TSS ne perturbe pas l'expression génique. Ce nombre de gènes insensibles est déterminé en fonction du seuil fixé pour dire que l'allongement de l'antisens conduit à une diminution de l'ARNm correspondant. Ici avec un seuil de répression très stringent de 2, il y a donc 27 gènes qui sont insensibles au phénomène d'interférence transcriptionnelle.

En résumé, ce projet a soulevé plusieurs questions quant au mécanisme d'interférence transcriptionnelle. Tout d'abord, les critères de sensibilité à l'interférence transcriptionnelle sont encore méconnus. Plus particulièrement, il nous a semblé intéressant de voir si cette sensibilité à l'interférence transcriptionnelle pouvait être déterminée par des propriétés intrinsèques aux promoteurs du gène régulé et/ou de l'antisens. D'autre part, ces propriétés des promoteurs sens et antisens correspondant aux gènes sensibles au phénomène pourraient potentiellement expliquer la spécificité de voie de modification de la chromatine impliquée dans le mécanisme de répression. Par ailleurs, l'identification de facteurs impliqués en *trans* dans ce mécanisme permettrait de mettre en évidence de nouvelles voies de régulation des gènes par interférence transcriptionnelle.

2. Développement d'un système rapporteur pour l'étude des mécanismes d'interférence transcriptionnelle

Pour répondre aux problématiques soulevées par les résultats précédents, ma thèse a d'abord consisté à essayer de mettre au point un système permettant de reproduire le phénomène d'interférence transcriptionnelle au niveau d'un gène rapporteur. L'idée était en effet de pouvoir quantifier l'expression d'un gène régulé par interférence transcriptionnelle sans obligation de mesure de l'abondance de l'ARNm, pour réaliser des tests phénotypiques après association du gène rapporteur à différentes combinaisons de promoteurs géniques et cryptiques. Nous souhaitions également utiliser le système pour réaliser des cribles génétiques.

a. Système rapporteur CUP1

Le premier des deux systèmes rapporteurs construit est une association du gène *CUP1* à une combinaison de promoteurs sens et antisens différents, afin de pouvoir observer les

conséquences de phénomènes d'interférence transcriptionnelle sur la résistance cellulaire au cuivre. Ce système présente l'avantage de permettre une réponse phénotypique graduelle aux régulations transcriptionnelles en jouant sur la concentration en cuivre dans le milieu, et donc de mieux détecter de faibles variations du niveau d'expression du gène rapporteur. Pour s'affranchir d'une déplétion de Nrd1 qui est néfaste pour les cellules, la copie du gène *CUP1* du système rapporteur a été mutée pour supprimer les sites de terminaisons NNS en antisens. Cette construction peut de plus être intégrée dans le génome au niveau du locus CUP.

Afin de vérifier que cette construction permette de reproduire et d'étudier le phénomène d'interférence transcriptionnelle, j'ai associé ce système rapporteur aux promoteurs sens et antisens d'un gène connu pour sa régulation par ce phénomène, le gène CPS1. Ce gène codant pour une carboxypeptidase vacuolaire est associé à un CUT antisens initié en 3'. En l'absence de NRD1, l'allongement de son transcrit antisens au-delà du TSS de *CPS1* conduit à une forte répression de l'expression du gène (Figure 17).

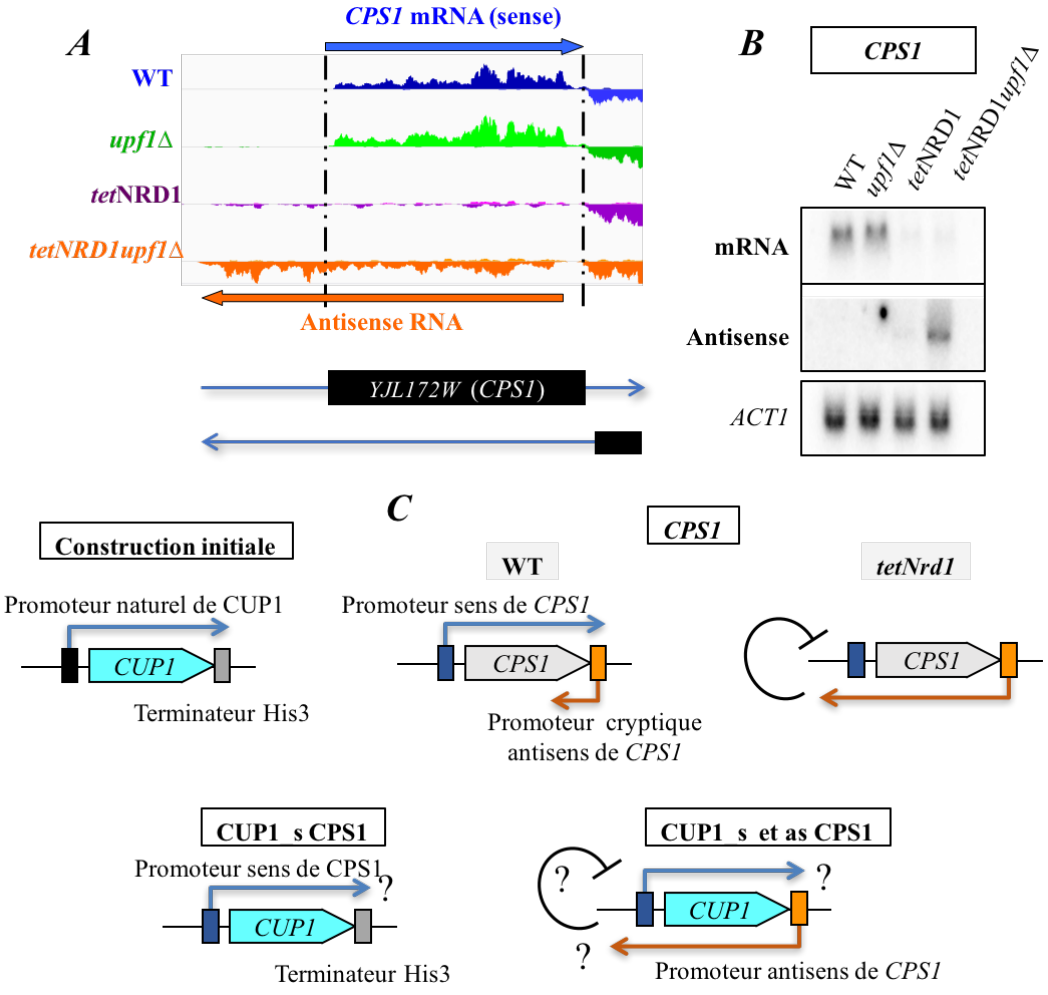


Figure 17 : (A) Représentation des résultats de RNA-seq. (B) Northern Blot montrant la répression du gène CPS1 par la transcription antisens *(Résultat non publié, Feurbach F., Malabat C. et Jacquier A.)*
(C) Schéma des associations du rapporteur CUP1 avec les promoteurs sens et antisens de CPS1

J'ai étudié l'effet de l'association du gène *CUP1* aux promoteurs sens et antisens de *CPS1* par des tests de croissance en milieux liquides et solides à différentes concentrations de cuivre. J'ai dans un premier temps cherché une différence de croissance entre une souche *cup1*Δ n'exprimant pas du tout le gène *CUP1* et une dans laquelle le système rapporteur contenant *CUP1* sous contrôle du promoteur génique de *CPS1* a été intégré (CUP1_s_CPS1). Vraisemblablement à cause d'un manque de sensibilité du système, la souche CUP1_s_CPS1 ne présente pas de résistance au cuivre supérieure à la souche *cup1*Δ, pour des concentrations en $CuSO_4$ allant de 0 à 80µM (Figure 18).

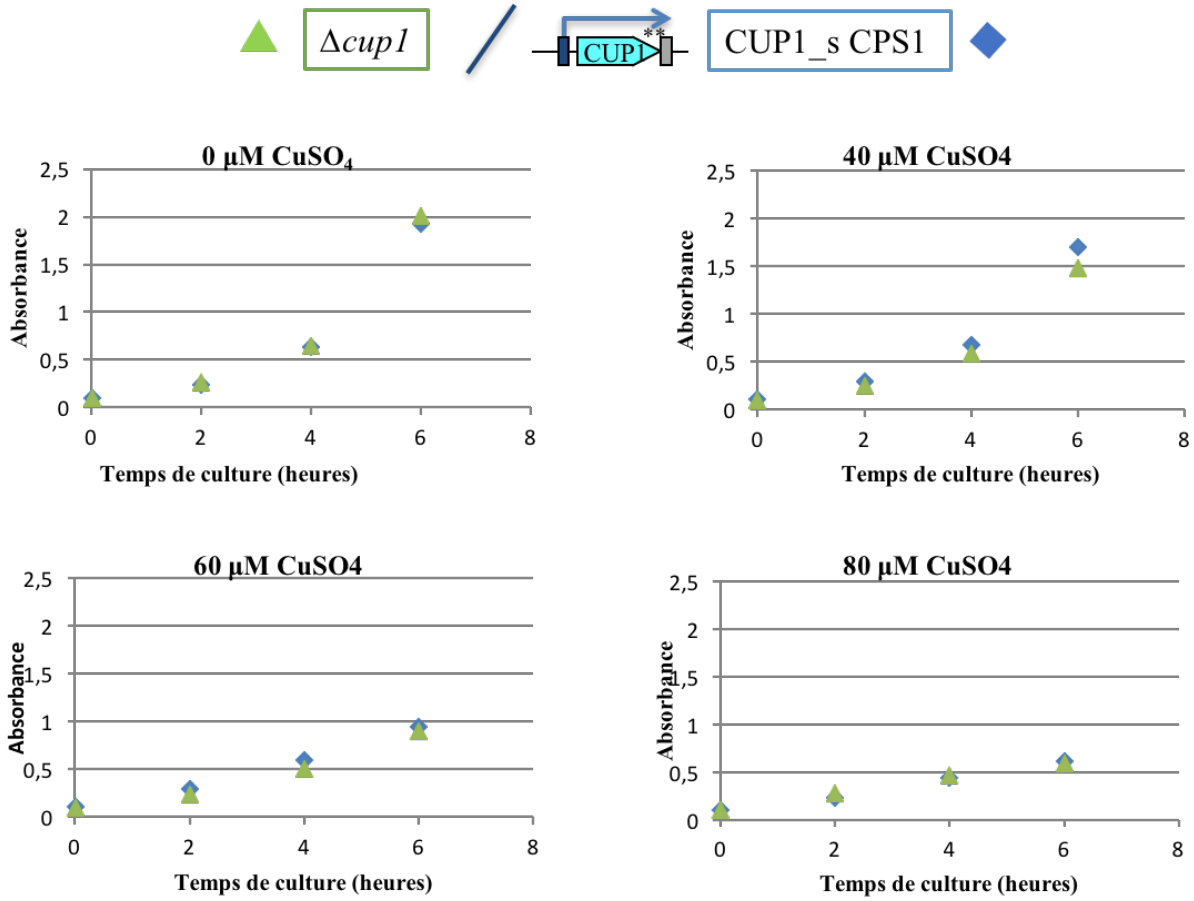


Figure 18 : Test de croissance différentielle en milieu liquide

J'ai ensuite réalisé des tests de compétition de croissance afin de voir si malgré la croissance apparemment identique des deux souches, une faible expression génique dans la souche CUP1_s_CPS1 lui permettait d'être néanmoins enrichie par rapport à la souche *cup1*Δ, au cours de plusieurs générations et en présence de cuivre. Pour cela, j'ai inoculé un milieu de

culture contenant 20 µM de $CuSO_4$, avec 99% de cellules *cup1*Δ et 1% de cellules CUP1_s_CPS1 (pour une absorbance finale de 0,05). La culture a été effectuée en turbidostat de manière à être maintenue en phase exponentielle par des dilutions régulières et automatiques à 30°C. Plusieurs prélèvements ont été effectués pendant environ 40 générations. La présence d'un marqueur de résistance à la Nourséothricine (NAT) dans le système rapporteur CUP (voir Matériel et Méthodes) a permis de suivre au cours des générations l'évolution du pourcentage de cellules CUP1_s_CPS1. Pour cela, les prélèvements effectués à des temps différents ont été étalés sur YPD, puis deux répliques sur YPD ou milieu [+ NAT] ont permis de décompter et comparer le nombre de cellules. En moins de 40 générations, l'augmentation du pourcentage de cellules CUP1_s_CPS1 de 1% à 40% montre que le gène CUP1 est bien exprimé lorsqu'il est mis sous contrôle du promoteur de CPS1 (Figure 19), bien que cette expression soit visiblement faible.

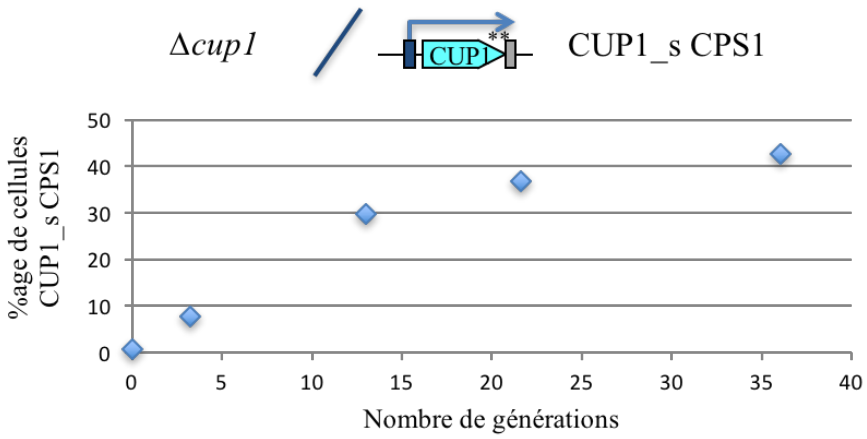


Figure 19 : Evolution du pourcentage de cellules CUP1_s_CPS1 au cours du test de compétition de croissance entre une souche cup1Δ et la souche CUP1_s_CPS1 à 20µm de CuSO4

La seconde étape consiste à vérifier que l'on peut reproduire le phénomène d'interférence transcriptionnelle, c'est à dire si l'on peut réprimer le gène CUP1 et donc diminuer la résistance cellulaire au cuivre en associant le rapporteur au promoteur antisens de CPS1. La construction contenant à la fois les promoteurs sens et antisens de CPS1 a donc été intégrée dans le génome de *S. cerevisiae*, et le locus HIS a été modifié pour que la souche CUP1_s et as_CPS1 devienne [HIS^+] et puisse ainsi être aisément distinguée de la souche CUP1_s_CPS1 (par étalement sur milieu dépourvu d'histidine). Les mêmes expériences de compétition de croissance ont été menées pour voir si la souche CUP1_s_CPS1 avait un avantage de croissance sur la souche CUP1_s et as_CPS1, à différentes concentrations de cuivre (Figure 20).

Les tests de compétition entre les souches CUP1_s_CPS1 et CUP1_s et as_CPS1 n'ont pas permis de conclure à la reproduction du phénomène d'interférence observé pour CPS1, au niveau du système rapporteur. De plus, un Northern-Blot réalisé à partir des ARN extraits des différentes souches utilisées, cultivées en milieu riche ou en milieu synthétique complet, ne permet pas de détecter un antisens du gène *CUP1* (Figure 21). En effet, le peu de signal que l'on observe pour la sonde antisens correspond en fait au signal de la sonde sens hybridée en premier. Ce système a donc été abandonné pour l'étude du mécanisme d'interférence transcriptionnelle.

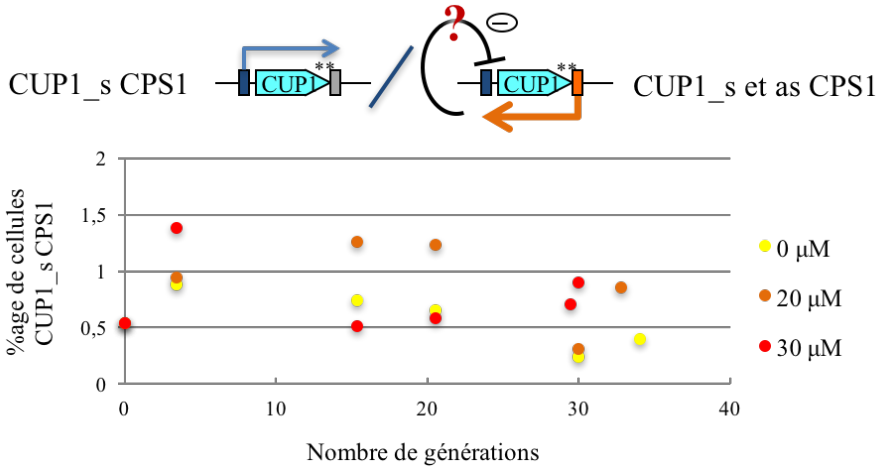


Figure 20 : Evolution du pourcentage de cellules CUP1_s_CPS1 au cours du test de compétition de croissance entre une souche CUP1_s_CPS1 et la souche CUP1_s et as_CPS1 à différentes concentrations de CuSO4

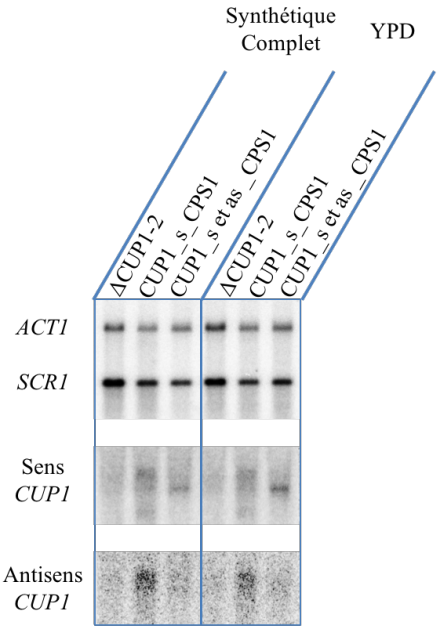


Figure 21 : Northern-Blot pour la détection des transcrits sens et antisens de CUP1, *au sein du système rapporteur associé à différentes combinaisons de promoteurs*

b. Système rapporteur URA3-GFP-mCherry

Le deuxième système rapporteur que j'ai testé utilise la fusion de trois phases de lecture, avec dans l'orientation sens les gènes *URA3* et *GFP*, et *mCherry* dans l'orientation antisens. Cette construction devrait permettre la mesure au niveau phénotypique à la fois de l'expression en sens et en antisens, et leur étude en cellules individualisées par microscopie. Le système rapporteur *URA-GFP-AS_mCherry* a été construit sur un plasmide centromérique pRS313, afin de limiter les différences d'expression observées entre cellules individuelles qui seraient dues à un nombre de copies différent du système rapporteur. La séquence codant pour la protéine mCherry étant presque dépourvue de site de terminaison NNS, nous avons considéré que la déplétion de Nrd1 n'était pas nécessaire à l'allongement de l'antisens.

De la même manière que pour le rapporteur *CUP1*, ce système a été associé aux promoteurs sens et/ou antisens de *CPS1* pour vérifier que chacun des transcrits était produit, et pour voir si cette association permettait de reproduire le phénomène d'interférence transcriptionnelle.
La construction initiale *URA3-GFP- AS mCherry* associée au promoteur du gène *ADE2* et au terminateur *HIS3,* permet bien une expression en sens détectée par une croissance sur milieu dépourvu d'uracile et par une expression de la *GFP* (Figure 22, panel du haut). Par ailleurs, l'association simple de la séquence codant la *mCherry* avec le promoteur antisens de *CPS1* et le terminateur *TEF* permet aussi la détection de la protéine fluorescente en microscopie (Figure 22, panel du milieu). Enfin, la fusion *URA3-GFP* est bien exprimée lorsqu'elle est placée sous le contrôle du promoteur de *CPS1* (Figure 22, panel du bas).

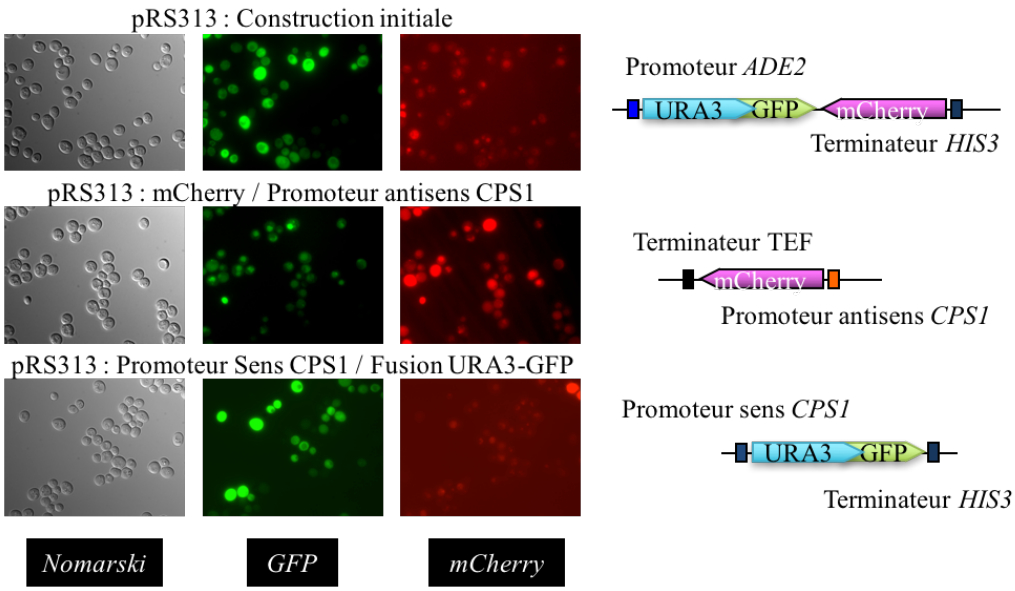


Figure 22 : Test de l'expression des gènes rapporteurs de la construction URA3-GFP-mCherry

Malgré ces premières vérifications, la construction du système rapporteur complet associée aux promoteurs sens et antisens de *CPS1* (URA3-GFP- *AS mCherry*_s et as CPS1) n'a permis de détecter aucune des deux protéines fluorescentes (Figure 23). Cela laisse penser que l'un et l'autre des transcrits sens et antisens pourraient être ciblés par les contrôles qualité cellulaires, tel que le NMD en raison de leurs longues séquences 3'-UTR, conduisant à leur dégradation et donc à l'absence de traduction des protéines fluorescentes.

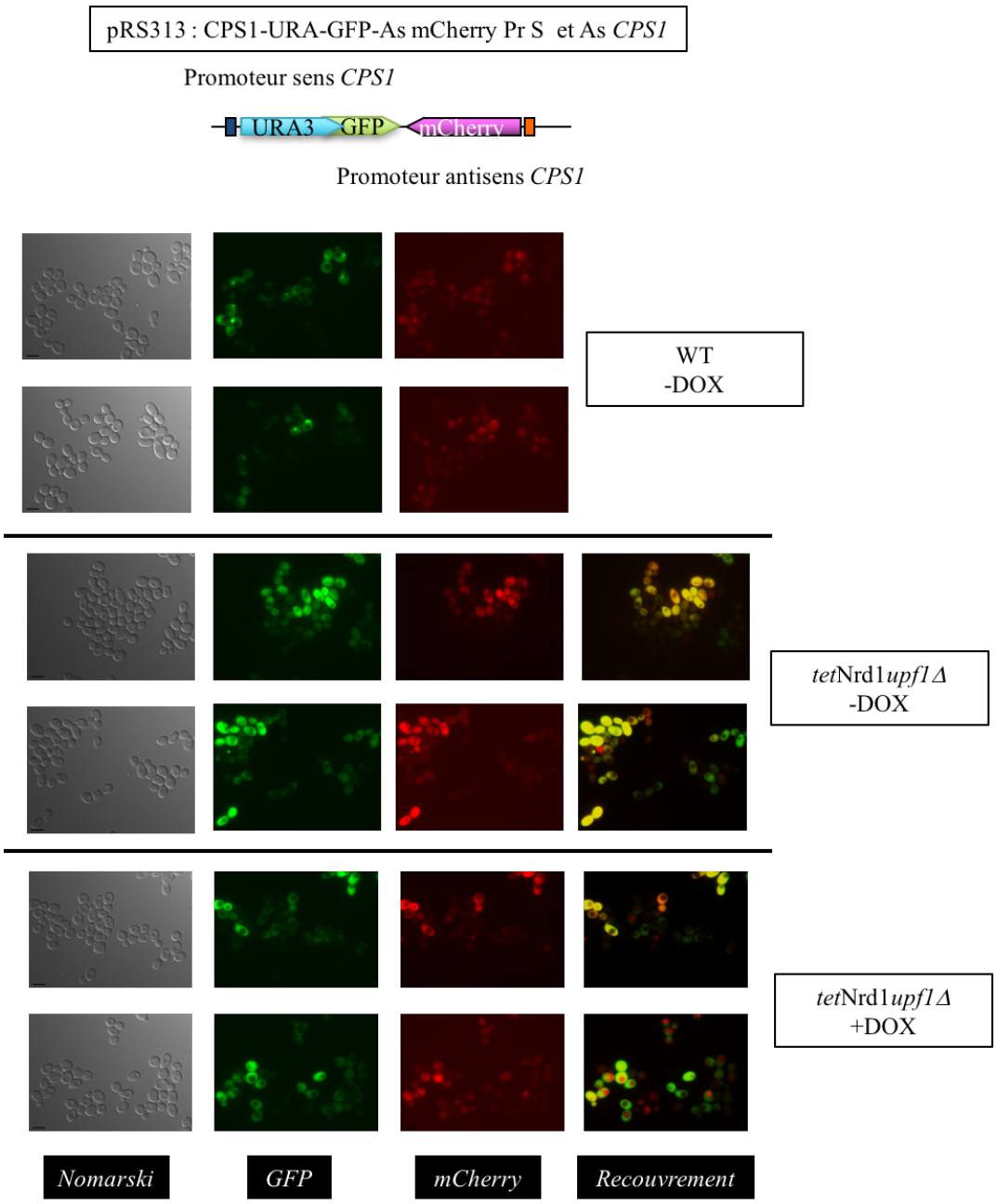


Figure 23 : Test de l'expression sens et antisens du système rapporteur

C'est pourquoi le plasmide contenant le système rapporteur a été transformé dans des cellules t*etNrd1upf1*Δ pour inhiber la dégradation des ARN par la voie NMD. La présence de doxycycline dans le milieu, qui réprime l'expression de *NRD1* dans ce fond génétique, n'a pas

d'effet sur l'expression de la *mCherry*, ni de la GFP, ce qui confirme que l'expression de ces fusions protéiques n'est pas restreinte par la voie NNS. En revanche, la GFP et la *mCherry* sont détectées dans les cellules suite à la délétion *upf1*Δ, ce qui confirme que les transcrits sont ciblés par une dégradation cytoplasmique *via* la voie du NMD. On remarque cependant que le taux d'expression de la GFP en présence de l'expression de l'antisens *AS mCherry* reste très élevé. De plus on observe qu'au niveau protéique, aucune occurrence d'expression mutuellement exclusive n'a lieu entre la GFP et la *mCherry* (Figure 23). Il nous a donc semblé que ce système n'était pas optimal pour l'étude en cellules individualisées des mécanismes d'interférence transcriptionnelle et ne permettait pas de réaliser les tests fonctionnels désirés.

II. Rôle de phénomènes d'interférence transcriptionnelle liés à la transcription antisens dans la régulation naturelle de l'expression génique (Manuscrit de l'article)

Compte tenu des difficultés rencontrées pour le développement d'un système rapporteur adapté pour l'étude des mécanismes d'interférence transcriptionnelle, je me suis orientée, dans une deuxième partie de ma thèse, sur l'étude d'exemples de régulations naturelles de l'expression génique par des phénomènes d'interférence transcriptionnelle liés à la transcription antisens non-codante.

Dans le but de rechercher de façon exhaustive les gènes ciblés par ces régulations, j'ai étudié le transcriptome de *Saccharomyces cerevisiae* et plus particulièrement la régulation des transcriptions génique et pervasive pendant la transition entre deux conditions environnementales. Cette analyse dynamique présente l'avantage de pouvoir suivre les régulations d'un type de transcription, ainsi que les conséquences potentielles sur l'autre.

J'ai ainsi analysé les régulations transcriptionnelles liées à la transition de cellules en phase exponentielle vers l'état de quiescence (G0). Chez *S. cerevisiae,* les cellules quittent le cycle cellulaire pour rejoindre la G0 après un épuisement en nutriments dans le milieu de culture. Elles subissent alors de nombreuses modifications morphologiques et métaboliques qui requièrent une reprogrammation complète de l'expression des gènes, notamment par des modifications transcriptionnelles (Coller *et al*., *Plos Biol.*, 2006 ; Gray *et al*., *Microbiol. Mol. Biol. Rev.*, 2004 ; De Virgilio, *FEMS Microbiol. Rev.*, 2012). Les gènes induits lors du passage en quiescence sont donc spécifiques de conditions de croissance particulières, et sont

peu exprimés pendant la croissance exponentielle. Nous avons notamment pu observer que ces gènes peu exprimés en phase exponentielle, sont associés dans cette condition à des ARN antisens fortement exprimés, potentiellement interférents pour la majorité d'entre eux car ils vont au-delà du TSS du gène associé. Des tests fonctionnels d'interruption de la transcription antisens pour quelques exemples nous ont permis de valider leur rôle fonctionnel en *cis* pour la répression transcriptionnelle. Des interruptions des transcriptions des ARNm correspondants, au moment où ces gènes sont normalement induits, nous ont également permis de mettre en évidence un rôle réciproque de la transcription génique qui réprime, en condition d'induction des gènes, la transcription non-codante antisens correspondante.

L'ensemble des résultats en rapport avec ce projet fait l'objet d'un article qui a été soumis pour publication fin Août, et qui est disponible sur bioRxiv.
(http://www.biorxiv.org/content/early/2017/07/28/169730).

Manuscrit de l'article de thèse

Antisense transcriptional interference mediates condition-specific gene repression in budding yeast

Title

Antisense transcriptional interference mediates condition-specific gene repression in budding yeast

Alicia Nevers[1,3], Antonia Doyen[1], Christophe Malabat[1,4], Bertrand Néron[4], Thomas Kergrohen[1], Alain Jacquier*[1,2] and Gwenael Badis*[1,2].

[1]Unité GIM, Institut Pasteur, Paris, France
[2]CNRS UMR3525, Paris, France
[3]Université Pierre et Marie Curie, Paris, France
[4] Bioinformatics and Biostatistics Hub C3BI, Institut Pasteur, Paris, France
*Corresponding authors:
Gwenael Badis Tel: +33 140613331; Fax: +33 140613456; email: gbreard@pasteur.fr
Alain Jacquier Tel: +33 140613205; Fax: +33 140613456; email: jacquier@pasteur.fr
Corresponding Author upon submission.
Gwenael Badis Tel: +33 140613331; Fax: +33 140613456; E‐mail: gbreard@pasteur.fr

Running title

Gene repression by antisense transcription

Keywords

Transcriptional interference/ antisense/ pervasive transcription/Quiescence

(54 885 characters).

Abstract

Pervasive transcription generates many unstable non-coding transcripts. Although a few examples of pervasive antisense have been shown to mediate gene regulation by transcriptional interference, whether pervasive transcription has a general functional role or merely represents transcriptional noise remains unclear. In a mutant context that stabilised pervasive transcripts, we characterized more than 800 antisense RNAs genome wide and

analysed the corresponding sense mRNA behaviour. We observed that antisense non-coding transcripts were associated with genes tightly repressed during exponential growth compared to quiescence and with an opposite level of variation between the mRNA and the associated antisense. This suggested that antisense transcription might participate to gene repression during the exponential phase. We thus specifically interrupted the antisense transcription of a subset of genes, and found that it resulted in a de-repression of the corresponding mRNAs. Antisense-mediated repression involved a cis-acting mechanism and was dependent on several chromatin modification factors. Our data convey that transcriptional interference by pervasive antisense transcription is a general mechanism of gene repression between cellular states.

Introduction

In steady state, the transcriptome reflects the equilibrium between RNA synthesis and degradation. Eukaryotes have developed sophisticated systems to control the turnover of mRNAs and ncRNAs necessary to the cell, undesired RNA species being rapidly eliminated by quality control mechanisms.

The development of genome-wide techniques to analyse transcriptomes, tilling arrays and cDNA next-generation sequencing, revealed that eukaryotic genomes are pervasively transcribed (Carninci & Hayashizaki, 2007). The genome of budding yeast is particularly compact and it has been hitherto conceded that more than 70% of it is composed of protein coding ORFs (Goffeau et al, 1996). Yet, this is only true if one does not distinguish the two DNA strands. If one takes into account sense and antisense sequences, non protein-coding genomic sequences represent up to 65% of the genome, leaving room to a large fraction of the genome for the generation of pervasive non-coding transcripts.

In yeast, pervasive transcription has been first reported more than a decade ago. If a fraction of it was uncovered in wild-type cells (David et al, 2006; Xu et al, 2009), a substantial part of the eukaryotic pervasive transcription is “hidden” as it generates very short-lived “cryptic” transcripts. These RNAs are difficult to detect unless they are stabilized by interfering with quality control mechanisms that normally eliminate them (Jensen et al, 2013). Pervasive transcripts detected in wild type yeast cells have been named “SUTs” for “Stable Unannotated Transcripts” (Xu et al, 2009), and different names have been given to cryptic transcripts depending on which factor was mutated in order to stabilise a particular class of RNAs. For example, CUTs, or Cryptic Unstable Transcripts were characterized upon removal of the exonuclease Rrp6, specific for the nuclear form of the exosome (Wyers et al, 2005; Neil et al,

2009; Xu et al, 2009), XUTs were revealed upon removal of the cytoplasmic exonuclease Xrn1 (van Dijk et al, 2011) and the NUTs correspond to the transcripts that accumulate when the nuclear termination factor Nrd1 is depleted (Schulz et al, 2013). Yet, there are in yeast only two main pathways responsible for the efficient elimination of pervasive transcripts, the nuclear Nrd1-Nab3-Sen1 (NNS) pathway, in which the early transcription termination of cryptic transcripts by the NNS complex is coupled to the degradation by the nuclear TRAMP-exosome complex (Arigo et al, 2006; Thiebaut et al, 2006; Schulz et al, 2013; Tudek et al, 2014) and the cytoplasmic non-sense mediated mRNA decay (NMD) pathway (Malabat et al, 2015; Wery et al, 2016). Many of pervasive transcripts require both pathways for their efficient and fast elimination (see Malabat et al, 2015).

Irrespective of which pathway predominates for their degradation, these transcripts all originate from nucleosome free regions (NFRs), which are essentially found 5' and 3' of mRNA coding sequences (Jiang & Pugh, 2009). When they originate from 5' NFRs, they are most often transcribed divergently from mRNAs and result from an intrinsic low polarity of gene promoters (Neil et al, 2009; Xu et al, 2009). This divergent transcription has the potential to interfere with the expression of the neighbouring upstream gene. Likewise, when a non-coding transcript initiates from the 3' NFR in antisense to the upstream gene, its transcription has the potential to interfere with the proper expression of the corresponding mRNA (van Dijk et al, 2011; Murray et al, 2012). Such transcription interference by pervasive transcription is largely prevented genome-wide by the NNS quality control pathway, which ensures the early transcription termination of these transcripts and prevent them to extend into the promoter region of the corresponding antisense genes (Arigo et al, 2006; Thiebaut et al, 2006; Schulz et al, 2013; Castelnuovo et al, 2014).

Whether pervasive transcription has a general function is a matter of debate. The fact that highly efficient quality control mechanisms have been selected during evolution to eliminate most of these transcripts argue in favour of the idea that most of them are non functional, but pervasive transcription itself, more than it product, could be important. Yet, the existence of the NNS pathway, which, by terminating pervasive transcription early, is key in preserving pervasive transcription from interfering with the expression of many coding genes, also suggests that a large fraction of these events simply result from the low specificity of RNA polymerase II (PolII) transcription initiation.

There are a number of well-documented examples of individual coding gene regulation through the transcription of a non-coding RNA: SER3 (Martens et al, 2004), IME1 and IME4

(van Werven et al, 2012), GAL10/GAL1 (Houseley et al, 2008; Pinskaya et al, 2009), PHO84 (Castelnuovo et al, 2013), CDC28 (Nadal-Ribelles et al, 2014). In the vast majority of cases analysed in budding yeast, the synthesis of a non-coding transcript has only an effect in cis. The prevailing model is thus that repressive chromatin marks are deposited in the promoter regions of genes in the wake of PolII transcribing the associated non-coding RNAs (Castelnuovo & Stutz, 2015; Murray et al, 2015). It is thus the act of transcription rather that its product, which is important. Several distinct mechanisms can be at play, but the general theme is that methyltransferases carried on the carboxy-terminal domain (CTD) of the PolII large subunit deposit histone methylation marks that recruit repressive chromatin modifiers such as histone deacetylases or nucleosome remodelling complexes. In budding yeast, there are two such CTD associated histone methyl transferases, Set1, which methylates histone H3K4 at promoters and gene proximal regions of actively transcribed genes, and Set2, which methylates H3K36 within more distal gene regions. The role of Set1 is complex. It is responsible for both H3K4 di- and tri-methylation (H3K4me2 and H3K4me3). It has been proposed that H3K4me3 at the beginning of actively transcribed genes could enhance and help maintaining preinitiation complex assembly and an active acetylated chromatin state, thus playing a positive role on transcription. Conversely, Set1 generates H3K4me2 in the body of gene, which recruits the histone deacetylase complexes SET3 or RPD3L, which repress transcription initiation (see Venters & Pugh, 2009 for review). Set2 is responsible for the H3K36 methylation (H3K36me2) in the body of genes, which results in the recruiting the Rpd3S deacetylase complex that plays an essential role in preventing improper internal initiation (Carrozza et al, 2005; Keogh et al, 2005). Thus both Set1 and Set2 have the potential to mediate transcription interference and have been implicated in gene repression by non-coding RNA transcriptional interference (see Castelnuovo & Stutz, 2015 for review).

How widely the transcription of non-coding RNAs, in particular antisense RNAs (asRNAs), is associated to gene regulation? Is it only restricted to a few exceptions, pervasive transcription then being merely the result of transcriptional noise, or does it play a broader role in gene regulation?

Several large-scale studies attempted to find a role to this transcription. Genes with large expression variability (such as stress response and environment specific genes) often have antisense expression and this suggests a general regulatory effect of antisense gene expression (Xu et al, 2011). Others correlated antisense expression with chromatin marks, either in a wild type context or in a rrp6 mutant condition (Murray et al, 2015; McKnight et al, 2015;

Castelnuovo et al, 2014; Gu et al, 2015; Kim et al, 2016) but no global anti-correlated trend was found between asRNA and mRNA expression.

Very recently, the question of the actual extent of the effect of asRNA transcription on gene regulation was addressed directly by measuring, at a large scale and under various conditions, the effect of specific antisense SUTs transcription interruption on the expression of the corresponding proteins fused to GFP (Huber et al, 2016). This study showed that, for 41 out of 152 (~27%) genes associated with an antisense SUT, a detectable antisense-dependent gene regulation could be observed under at least one condition. Although no specific biological pathway seemed enriched in the asRNA responsive genes, the analysis showed that repression by asRNA transcription interference helps reducing mRNA expression basal levels, especially for genes expressed at a low level, resulting in complete gene repression.

This analysis is restricted to SUTs, i.e. non-coding RNAs readily detectable in wild-type cells. Yet, we know that SUTs represent only a minority of the pervasive transcripts, most of which are too unstable to be detected in wild-type cells (Neil et al, 2009; Xu et al, 2009; van Dijk et al, 2011; Geisler et al, 2012). Whilst the nuclear NNS quality control pathway terminates early the transcription of many of the pervasive RNAs, to prevent them from interfering with mRNA expression (Vasiljeva et al, 2008; Schulz et al, 2013), many pervasive transcripts escape, at least in part, this first surveillance pathway and are extended up to cryptic cleavage and polyadenylation sites (polyA sites), potentially over the transcription start site (TSS) of their associated genes. This can lead to the export of long non-coding RNAs into the cytoplasm, where they are rapidly degraded by the NMD pathway (Malabat et al, 2015; Wery et al, 2016).

We analysed genome-wide the amount of asRNAs associated to each mRNA in a NMD mutant context (upf1Δ). In this mutant, hidden pervasive transcripts that escaped the nuclear NNS surveillance accumulate in the cytoplasm and can thus be quantified. An important fraction of the less expressed genes are associated with asRNAs, especially if the asRNAs overlapped the associated sense gene promoter. In addition, many of these genes with promoter-overlapping asRNAs were enriched for genes up-regulated in chromatin remodelling mutants such as set2Δ or set1Δ. Many of these mRNAs, poorly expressed during the exponential phase, corresponded to mRNAs found enriched during the stationary phase (G0). These observations strongly suggested that this particular class of genes is frequently subjected to asRNA transcription interference in order to be fully repressed during logarithmic growth, a prediction we validated for a subset of examples. This showed that

antisense-mediated transcriptional interference is a common mechanism used when mRNA expression needs to be tightly repressed under specific conditions.

Results

Characterization of antisense transcription in upf1Δ cells

In order to reveal antisense transcription that escaped the nuclear surveillance by the NNS pathway and reached the proximal region of the genes, we quantified the asRNAs levels over a +1 (mRNA TSS) to +200 nucleotides window of protein coding genes in a strain impaired for the cytoplasmic NMD surveillance pathway (upf1Δ mutant). Figure 1A shows these values (y axis) plotted against the average mRNA levels per gene (number of reads per nucleotide; x axis). The less expressed mRNAs appeared to be generally associated with high levels of asRNAs. In order to quantify this observation, we partitioned the genes according to their mRNAs levels in ten bins with an equal numbers of genes. The less expressed genes (bin 1) had significantly higher levels of asRNAs than the genes within higher mRNA expression categories (bins 2 to 10; see Figure 1B, Table EV1 and Dataset 1).

At least two non-exclusive phenomena could explain this observation. First, mRNA transcription itself could have a repressive effect on asRNA transcription initiation from their corresponding 3' NFRs (Xu et al, 2011). Hence, asRNAs initiating within NFRs situated downstream of non-expressed genes should be less subjected to such repression by mRNA transcription. Conversely, antisense-transcription from 3' NFRs could be a common mean to contribute to a tight gene repression. If the former explanation is correct, asRNAs associated with non-expressed genes should not show different termination characteristics than other asRNAs. In contrast, it was shown that repression by asRNAs correlates with mRNA TSS overlap (Huber et al, 2016). If asRNAs associated with the less expressed genes contribute to their tight repression, these asRNAs should overlap the mRNA TSSs more often than other asRNAs. We thus categorized asRNAs depending on their occurrence across TSSs by analysing a window between -50 nucleotides to +200 nucleotides relative to the mRNA TSS. We defined three types of genes. Genes without substantial asRNAs over this region defined class N (No antisense - arbitrarily set below three reads per base; see Dataset 1). Genes with asRNAs terminating before the mRNA TSS defined class M (mRNA antisense). Genes with TSS overlapping asRNAs defined class O (Overlapping antisense) (Figure 1C). Figure 1D shows a heat map of the sense and antisense transcripts over a -200 to +200 nucleotides

around the mRNA TSSs, ordered by these classes. Among the 5892 protein coding genes analysed, 5076 belong to class N, 259 to class M and 557 have an overlapping asRNA (class O). The higher proportion asRNAs in bin 1 mostly resulted from the over-representation of class O asRNAs (Figure 1E, Figure EV1 and Table EV2), strongly suggesting that many of them could be associated to antisense transcription mediated repression.

Genes up-regulated in the absence of RNA PolII associated chromatin regulators are enriched in the class of poorly expressed genes with TSS-overlapping asRNAs

If antisense transcription can affect sense transcription, one should expect that genes associated with asRNAs should be more up-regulated in chromatin modifier mutants implicated in transcriptional interference, in particular set1 and set2 mutants. Given that antisense transcriptional interference involves the extension of asRNA up to the promoter regions of the genes, Set2, which promotes H3K36me2 at late stages of PolII elongation, seemed a good candidate to mediate gene repression by asRNA transcription. Set2 mutants are intrinsically difficult to analyse by RNAseq or tilling arrays since one of its main role is to suppress internal initiation both in sense and antisense of gene transcribed regions (Carrozza et al, 2005; Keogh et al, 2005; Venkatesh et al, 2012; Malabat et al, 2015). The cryptic initiation events observed in set2Δ mutants in the sense orientation can thus induce misleading global increase of the mRNA quantification by contributing to the overall sense RNAseq counts (Malabat et al, 2015). We thus took advantage of the analysis of individual TSSs in the Malabat et al. study, which allows the analysis of the specific mRNA TSSs, irrespective of internal transcription initiation. We considered a gene as up-regulated upon SET2 deletion when its strongest mRNA-linked TSS cluster was induced at least two fold and with a p-value ≤ 0.05 (Supplementary file 3 in Malabat et al, 2015). Ninety-five of 5228 genes analysed in this dataset were up-regulated in set2Δ (see Dataset 1) Strikingly, genes with TSS-overlapping asRNAs (class O) showed the highest percentage of up regulation in set2Δ (8.3% compared to 1.8% for all genes; Figure 2A). Combining gene promoter classes with the mRNA expression level categories drastically increased this bias since class O of bin 1 showed the highest proportion of genes up-regulated in set2Δ cells (Figure 2B, right panel).
Direct measures of transcription levels by NETseq were also analysed in a set2Δ mutant (Churchman & Weissman, 2011). Although a higher number of genes were found up-regulated in absence of Set2, possibly due to internal initiation events not being filtered out, the same trend was observed (Figure EV2A). This prompted us to analyse the data for the

set1Δ, as well as rco1Δ and eaf3Δ (two components of the Rpd3S deacetylase complex) mutants from the same dataset, as these factors have also been found involved in transcription interference. Genes up-regulated upon deletion of these factors were also clearly over represented in class O (Figures EV2B-D). Altogether these results suggest that repression by antisense transcriptional interference is frequent for poorly expressed genes, and is mediated by several, chromatin-modifying factors linked to elongating PolII. Figure EV2E reports the overlapping number of up-regulated genes in the different mutant strains.

Quiescence enriched genes are associated with high levels of asRNAs

We next questioned whether poorly expressed genes (bin 1) could belong to a particular category of regulated genes. An expected category of genes strongly repressed during exponential growth are genes found enriched is stationary phase and/or in quiescence (G0). We thus analysed a dataset reporting the time course of mRNA expression over a complete 9-days growth cycle (Radonjic et al, 2005). Figure 3A shows that the stationary phase enriched genes (SP-enriched in Radonjic et al, 2005) were most abundant in bin 1. The Radonjic dataset essentially documents mRNAs expression in wild-type cells. In addition, stationary phase cell populations are not homogeneous, being composed of dead, dying and quiescent cells (Allen et al, 2006; Aragon et al, 2008). We thus decided to perform a genome-wide RNAseq analysis of quiescent (G0) cells, which represents a much more homogeneous cell population, in wild type or upf1Δ cells in order to analyse both gene and pervasive transcription. To normalise the overall level of transcripts per genome, we spiked in the budding yeast cultures before RNA extraction with identical reference aliquots of a Schizosaccharomyces pombe exponential culture (see Materials and Methods for normalisation procedures). We defined quiescence-enriched (Q-enriched) mRNAs as being, after normalisation, five times more abundant in the G0 population compared to the cells in exponential growing phase (260 genes; see Table EV3). As anticipated, Q-enriched mRNAs were mostly found in bin 1 (Figure 3B and C). Accordingly, Figure 3D shows that, as for genes within bin 1, Q-enriched mRNAs were associated with higher asRNA levels than random (*** with p = 7.4 10-5) and their distribution in the different asRNA associated genes classes (classes N, M & O) was similar to that of bin 1 (Figure EV3A). Moreover, while 8.3% of all genes with TSS-overlapping asRNAs were up-regulated upon deletion of SET2 (Figure 2A), this proportion raised to more than 35% when only considering Q-enriched genes (Figure EV3B). Breaking down these figures by bins and promoter classes showed that this

strikingly high proportion was primarily contributed by class O genes, representing 14 out of 22 (64%) of the set2Δ up-regulated Q-enriched genes (Figure 3E). These observations strongly suggested that asRNA transcriptional interference could be a frequent mechanism of tight repression for this class of genes. In order to directly test this hypothesis, we chose for further analyses five representative examples of Q-enriched genes associated with an asRNA spanning the mRNA TSS: PET10, SHH3, MOH1, CLD1 and ARO10. Among these genes, only ARO10 was previously tested (Huber et al, 2016) for asRNA mediated transcription interference.

Time course of quiescence enriched mRNAs and corresponding asRNAs show an inverse expression pattern

In order to examine the relative behaviour of these mRNAs and their associated asRNAs during a growth cycle, we performed Northern-blots time course experiments starting (t0') with addition of rich medium to quiescence purified cells and using strand-specific RNA probes. The five selected Q-enriched mRNAs were not only accumulating during quiescence but were in fact strongly induced after about 48 hours of culture (Figure 4), which coincides with the post diauxic shift transition (Radonjic et al, 2005). The asRNAs started to accumulate between 5 and 30 minutes upon rich medium addition, to reach a peak of expression at ~24 hours, after which they rapidly disappeared. The mRNAs followed the inverse trend, although ARO10 was not as substantially repressed during the exponential phase. These observations are compatible with the asRNA transcription contributing to mRNA repression. Conversely, they are also compatible with induction of the mRNA repressing the associated asRNAs.

Interruption of antisense transcription allows a de-repression of Q-enriched genes

One of the main effects of the NNS pathway is to prevent the expression of most pervasive transcription from interfering with the normal expression of genes genome-wide (Schulz et al, 2013). This mechanism is thus intrinsically optimized to terminate transcription early and in a strand specific manner. We thus chose to use it in order to specifically terminate asRNA transcription close to their transcription start by introducing in the TSS proximal region of the asRNAs a short (37 nucleotides) optimal NNS termination signal (NNS-ter; Porrua et al, 2012). In order to perturb as little as possible the corresponding mRNAs, we introduced this NNS-ter sequence seamlessly using a cloning-free method allowing chromosomal

modifications without leaving selection markers (Erdeniz et al, 1997). This sequence was introduced in the proximal region of the asRNAs, corresponding to the terminal region of the mRNAs (see Figure EV4 and Table EV4). The introduction of the NNS-ter signal resulted in the proper elimination of all asRNAs and in a strong up-regulation of the corresponding mRNAs, except for ARO10 (Figure 5). We note that ARO10 is also, out of the five genes examined, the one that showed the weakest mRNA repression during the exponential phase (see Figure 4 and Discussion).

The asRNA associated gene repression acts in cis

Although the majority of non-coding RNA associated gene regulation has been shown, when this was examined, to act only in cis, a trans effect of the asRNA itself has been invoked in a few cases (see Castelnuovo & Stutz, 2015 for a review). We directly addressed this question on PET10 by comparing the mRNA and the asRNA expression in cis and in trans. In order to carry out this experiment, we built diploid strains where PET10 sense and antisense transcripts were disrupted on one or two of the homologous chromosomes, allowing the expression of the asRNA either from the same chromosome as the mRNA (in cis), from the opposite chromosome (in trans) or no asRNA expression (see Figure 6A). RT-qPCR measurement showed that PET10 mRNA is repressed only when its asRNA is expressed in cis (blue) but not in trans (green). In this case the mRNA level reached the same level as observed in the control strain without antisense (red). This is fully consistent with the hypothesis that the antisense transcription and not the asRNA itself, acts to repress mRNA by a transcriptional interference mechanism.

Several PolII elongation-associated chromatin modification factors cooperate to mediate antisense transcriptional interference

As described above, TSS-overlapping asRNA associated genes were more prone to be up-regulated upon deletion of chromatin modifiers such as SET2 (Figure 2 and EV2) or SET1, RCO1 and EAF3 (Figure EV2, A-D) than the other categories of genes. Although the effects of set2Δ, rco1Δ and eaf3Δ are expected to be largely redundant as these factors act in the same chromatin modification pathway (Carrozza et al, 2005; Keogh et al, 2005; Churchman & Weissman, 2011), we also observed that more than half (80 out of the 155) of the genes that we computed in the Churchmann dataset (Churchman & Weissman, 2011) as the most

up-regulated in set2Δ were also up-regulated in set1Δ (Figure EV2 E). This suggested that these different chromatin modifiers might cooperate to mediate asRNA transcriptional gene repression. The interpretation of such data are complicated by the fact that chromatin modifiers can affect both the mRNAs and their associated antisense (Murray et al, 2015; Castelnuovo & Stutz, 2015). To address this question, we directly measured the effects of set1Δ, set2Δ and hda1Δ on both the asRNA and the mRNA by strand-specific RT-qPCR (see Material and Methods). The analysis of the PET10 locus indeed shows a complex picture. Not only the mRNAs were positively affected in different mutants, but the levels of asRNA was also impaired in all these mutants, making the evaluation of the antisense transcription interference on the mRNA difficult. In contrast, the SHH3 asRNA was not repressed upon these deletions while the mRNA was significantly derepressed in all the mutants, although not at the level of the control strain in which the asRNA transcription elongation is restricted by the NNS-terminator (Figure 6B). This suggests that several chromatin modification pathways cooperate to mediate an efficient transcriptional interference to repress gene expression.

Sense and antisense transcription can mutually repress each other

As discussed above, the fact that the category of genes associated with asRNAs was enriched in the least expressed genes (bin 1 or quiescence enriched genes) could result from two non-exclusive phenomena: either the asRNA transcription represses the mRNA or the absence of mRNA expression allows pervasive asRNA transcription from 3'-NFRs. We showed that, in four out of five genes we tested, asRNA transcription interruption led to an increase of sense mRNA levels, indicating a strong repressive effect of asRNA transcription on mRNA transcription. Indeed, as observed in Figure 4, the expression time courses of the mRNAs and asRNAs present inversed expression patterns, which is compatible with the mutual repression of sense and antisense transcription. We tested this hypothesis by interrupting the mRNA transcription of the PET10, SHH3, ARO10 and MOH1. Figure 7A shows that in all cases but for MOH1, restricting mRNA transcription elongation led to an up-regulation of the corresponding asRNA at 48 hours (post-diauxic shift). The absence of effect observed for MOH1 can be understood by the fact that its polyA site is the only of the four gene analysed, which is located upstream of its associated asRNA TSS (see Figure EV4). The mutual repression of sense and antisense transcription can thus be observed but will depend on locus-specific architecture.

As shown above, the mRNAs and their associated asRNAs exhibited an inverse pattern of expression between the exponential phase and quiescence, as expected if the expression of sense and antisense were mutually exclusive. Q-enriched genes were associated with more antisense transcription than average during the exponential phase, when these genes are repressed. Conversely, one could thus expect that they would be associated with less asRNA level than average in quiescence since they are the genes whose mRNAs are most abundant under this condition. This turned out not to be the case. Indeed, the Q-enriched genes remained associated with slightly higher asRNA levels than average even during quiescence (Figure 7B). This is consistent with the observation that there is no obligatory repression of asRNA transcription when sense transcription is induced and with the observation that the Q-enriched genes are, overall, more associated with asRNAs than other genes.

The analysis of sense and antisense expression in quiescence also revealed that, globally, if the mRNA levels strongly decreased in quiescence, as expected, the global level of asRNAs did not change (Figure EV5). A likely explanation is that mRNA transcription globally interferes with pervasive asRNA transcription during exponential phase. The global repression of transcription in quiescence (McKnight et al, 2015) could then be compensated for the asRNAs by the decrease of interference by mRNA transcription. This was verified at the HIS1 locus, where the strong asRNA observed only in quiescence (Figure EV6A) could be revealed during the exponential phase by interrupting the HIS1 mRNA transcription by a strand specific NNS terminator insertion (Figure EV6B).

Discussion

The discovery of an extensive pervasive transcription in yeast raised the question of whether it merely reflects transcription "noise" or whether it carries a widespread function. A large fraction of the yeast pervasive non-coding RNA transcription generates "cryptic" transcripts, which are rapidly degraded by the efficient NNS and NMD surveillance pathways. The NNS pathway restricts RNA PolII elongation of non-coding pervasive transcripts in order to prevent it from interfering with gene expression (Arigo et al, 2006; Thiebaut et al, 2006; Schulz et al, 2013). Yet, a very large number of pervasive transcription initiation events give rise to transcripts that escape, at least in part, this first nuclear surveillance pathway. These transcripts are then terminated by the cleavage and polyadenylation machinery and exported to the cytoplasm as regular capped and polyadenylated RNAs. These RNAs, while non-coding in the sense of coding for functional proteins, all carry spurious ORFs and are thus engaged in translation. These short spurious ORFs are followed by long non-coding 3'-UTR

like sequences. These transcripts thus exhibit all the characteristic features of efficient NMD targets and are therefore rapidly degraded by this cytoplasmic pathway (Malabat et al, 2015). Many of these transcripts are elongated from NFRs located at the 3'-end of genes. Those that are not fully terminated early by the NNS pathway can be elongated up to the promoter region of the corresponding gene, potentially interfering with their expression.

Targeted studies have previously described specific examples in which the transcription of a non-coding RNA was mediating gene regulation (see for reviews Donaldson & Saville, 2012; Castelnuovo & Stutz, 2015). To what extent this phenomenon results in a significant repression of the corresponding genes genome-wide is a question that has been recently addressed by Huber and colleagues, who performed a large-scale study focusing on the SUTs. This analysis concluded that less than 25% of the 162 mRNA/asRNA pairs analysed showed a weak suppressive effect of the transcription of antisense SUTs on the expression of the corresponding protein, in at least one condition (Huber et al, 2016). Antisense transcripts overlapping the mRNA TSS of weakly expressed genes were the most likely to exert such a suppressive effect.

This analysis, whatever comprehensive, is nevertheless restricted to a relatively small fraction of the naturally occurring antisense transcription since it focused only on SUTs. Analysing the transcriptome of NMD deficient cells (upf1Δ), we classified 816 genes as being associated with asRNAs (Figure 1). It was previously reported that the levels of antisense transcription were not correlated with that of the corresponding mRNAs (Murray et al, 2015). Figure 1A plots the respective levels of asRNAs and corresponding mRNAs in upf1Δ cells. While, similar to the Murray analysis, a linear regression analysis did not reveal any correlation between the two datasets (Pearson correlation coefficient R2=0.05), we nevertheless observed that many of the least expressed genes were associated with high antisense levels. Dividing the 5892 genes of the dataset in ten bins of equal size, we found that the least expressed genes, corresponding to bin1, were significantly more associated with asRNAs than the other genes (Figure 1B). This could result in part from the fact that, in contrast to well-expressed genes, the asRNA transcription of the least expressed genes is not repressed by sense transcription. Yet, almost 78% (181/233) of these asRNAs overlapped the mRNA TSS of their corresponding genes in bin 1 and this percentage was very significantly higher than in other bins (Figures 1E, EV1 and Table EV2). Since this class of asRNAs has been shown to be more prone to suppress gene expression when compared to non TSS-overlapping asRNAs (Huber et al, 2016), it suggested that these least expressed genes could frequently be repressed by asRNA transcription interference. This idea was further strengthened by the observation

that TSS overlapping asRNA associated genes are much more frequently up-regulated upon SET2 deletion than average (Figure 2A) and this biais is for a large part contributed by genes in bin 1 (Figure 2B). Other chromatin modification factors, such as SET1, RCO1 and EAF3 showed the same trends (Figure EV2). Interestingly, while asRNA transcription mediated regulation was found to affect different genes of very diverse pathways, we uncovered that genes whose mRNAs accumulated in quiescence (Q-enriched genes) were particularly enriched in bin 1 and in the class of genes associated with TSS-overlapping asRNAs (Figures 3 and EV3).). It suggested that the full repression of this class of genes during the exponential phase often relies on interference by antisense transcription. We directly tested this hypothesis by analysing five of these Q-enriched genes associated with TSS-overlapping asRNA (PET10, CLD1, MOH1, SHH3 and ARO10). For four out of these five genes, specifically interrupting asRNA transcription resulted in a strong induction of the corresponding mRNAs during the exponential phase (Figure 5). Interestingly, the only gene that did not respond was ARO10 but it was also, out of the five genes analysed, the least repressed gene during the exponential phase (Figure 4). In addition, this was the only gene we analysed that was also analysed in the Huber study (Huber et al, 2016). Although they could not find a repressive effect of its associated asRNA in rich medium, they found it to be regulated by antisense transcription when the cells were grown in synthetic complete medium. It thus turns out that the TSS-overlapping asRNAs associated to all five Q-enriched genes we tested can have a repressive role on gene transcription.

Interestingly, we found the repression of the asRNA upon induction of the mRNA to be frequent, although not obligatory and depending on the fact that the induced mRNA transcription overlaps the asRNA TSS (Figure 7 and EV4). Overall, the asRNA levels remained high in quiescence, even slightly higher than average (Figure 7B, right panels). It suggests a model by which, in contrast to previously studied examples (see for examples Martens et al, 2004; van Werven et al, 2012), the asRNA expression is not regulated by specific transcription regulators. Rather these RNAs would be constitutively expressed, unless repressed by sense transcription when mRNAs are induced and overlap their TSSs. Their transcription would thus act “passively” as an amplifier of gene regulation, turning an non-induction into repression, as previously suggested for the SUR7 gene (Xu et al, 2011). Consistent with this model, blocking asRNA transcription elongation during the exponential phase resulted in a 2.6 and 8.3 fold increase of PET10 and SHH3 mRNAs respectively (Figure 6B), which is markedly lower than the induction estimated by comparing the increase of their relative expression levels measured from the quiescence versus exponential phase

transcriptome datasets (5.3 and 93.4 fold increase respectively in the upf1Δ background; dataset 1).

If, in a few instance, the asRNA itself was suggested to play a direct role in gene repression, in the majority of cases examined thus far this repressive effect was shown to be mediated in cis by antisense transcription interference, the asRNA being only a by-product of this process (see for review Donaldson & Saville, 2012). In the case where the asRNA itself is proposed to play a direct effector role, one of the main arguments in favour of this model is that the repression is observed only when the non-coding RNA is stabilised by mutating either the nuclear exosome (Berretta et al, 2008) or the cytoplasmic Xrn1 exonuclease (van Dijk et al, 2011). This is not the case for the examples we studied, the repressive effect being identical in a wild type or in a NMD deficient background, in which the asRNAs were stabilized. Using strand specific NNS terminators in a diploid strain, we directly confirmed, on the PET10 locus, the prediction that the effects we observed act only in cis, which is a hallmark of the transcription interference mechanism (Figure 6A).

Interestingly, the presence of a PET10 asRNA with an inversed expression profile compared to the mRNA was shown to be conserved in all five analysed Saccharomyces species, supporting its functional role (Yassour et al, 2010). More generally, a phylogenetic conservation study of lncRNAs in budding yeasts has shown that, since the divergence with N. castellii, which has retained a functional RNAi machinery, the level of asRNAs and their extent has globally increased. It was thus suggested that the lack of RNAi favoured the development of asRNA transcription mediated gene regulation (Alcid & Tsukiyama, 2016).

Transcriptional interference was previously shown in yeast to rely on the chromatin modifications by factors bound to the elongating PolII (for review Castelnuovo & Stutz, 2015). The genes up-regulated in the absence of factors such as SET2, SET1, RCO1 or EAF3 were found more often associated to TSS-overlapping asRNA than average (Figures 2, EV2), and this was particularly true for Q-enriched genes (Figures 3E and EV3B). The specific study of the SHH3 gene by RT-qPCR showed that the suppressive effect of the antisense transcription involved a combination of several factors, acting in different chromatin modification pathways (SET2, SET1, HDA1; Figure 6B). Taken together, these results strongly suggest that the observed asRNA-transcription mediated repression involves several redundant chromatin modification/remodelling pathways. This is reminiscent of previous

observations showing that gene silencing is mediated by redundant mechanisms involving multiple histone modifiers (Verzijlbergen et al, 2009). In agreement with this idea, each individual mutation could only promote a partial de-repression of the SHH3 gene when compared to the level observed in the positive control strain, in which the asRNA transcription is interrupted by the NNS terminator (AS NNS, Figure 6B). It suggests the possibility that any gene, associated to a TSS-overlapping asRNA, might be subjected to a regulation by a specific combination of chromatin modifiers.

In our study, we demonstrated that TSS-overlaping antisense-mediated transcriptional interference is a frequent mechanism used for full gene repression. This mechanism is often hidden since these antisense transcripts are rapidly degraded by the NMD pathway and therefore not detected in wild type conditions. We and others reported the existence of conditional asRNAs, such as for example the HIS1 antisense RNA, specifically expressed during G0 (Figure EV6A), or Meiotic Unannotated transcripts (MUTs; Lardenois et al, 2011). In addition, asRNAs were shown to mediate protein expression regulation depending on various growth conditions (Huber et al, 2016). Widespread antisense transcription has thus the potential to repress the synthesis of sense RNA and participate to differential gene expression and adaptation to various environmental and growth conditions.

Materials and Methods

Detailed strains, cultures, libraries preparation, Northern blot and RT qPCR methods are detailed in Appendix supplementary Methods.

Data analysis

Illumina reads treatments

For RNAseq libraries, duplicated reads were first filtered out using fqduplicate (ftp://ftp.pasteur.fr/pub/gensoft/projects/fqtools/fqtools-1.1.tar.gz). Then sequencing error were corrected using Musket (Liu et al, 2013; version 1.1). Reads of bad quality were removed using fastq_qual_trimmer (https://github.com/ivars-silamikelis/fastq_qual_trimmer, version 1.0) with a threshold of 20. Illumina adpaters were finally removed using Flexbar (Dodt et al, 2012; version 2.7). After removal of the random sequence tag, resulting reads were mapped using bowtie (Langmead & Salzberg, 2012; version 2.2.3 with the following parameters: –N 1 –p 1 —no-unal –D 15 –R 2 –L 22 –I S,1,1.15) and a compilation of S. cerevisiae genome (S288C reference sequence, Release 64 obtained from the Saccharomyces Genome Database (SGD) [http://www.yeastgenome.org/]) and S. pombe genome (ASM294 reference sequence, v2.19 obtained from PomBase [http://www.pombase.org/]) as reference genomes.

For 3'Long SAGE libraries, duplicated reads were first filtered out using fqduplicate. Illumina adpaters were then removed using AlienTrimmer (Criscuolo & Brisse, 2013). Reads corresponding the 3' end of transcripts were identified by detection of a polyA sequence at the end of the reads with a minimal size of 6 nucleotides. After Poly A removal, the resulting reads were mapped using bowtie (same version and parameters that above) and the S. cerevisiae genome (S288C reference sequence, Release 64 obtained from the Saccharomyces Genome Database (SGD) [http://www.yeastgenome.org/]). False positive reads (ie reads identified by a $\geq$ 6nt encoded polyA sequence but not a true 3'end) were filtered out by matching with encoded PolyA sequence in the genome.

Mapped reads processing

For 3’ Long SAGE libraries the 3′ -end positions of the resulting mapped reads were used as TTS positions and extracted to wig files. For RNAseq libraries, reads corresponding to the whole transcripts and full read coverage were extracted to wig files.

Normalisation and differential expression

Transcript differential expressions were calculated using DESeq2 (4) within the SARTools pipeline (Varet et al, 2016; version 1.4.1).

Sample corresponding to cells in exponential phase were first treated together as a separated group, as well as samples corresponding to cells in G0. During the process, SARTools performed a normalisation step. Normalisation factors were extracted and used to produce normalized wig files.

G0 samples were normalized in a second time against exponential phase sample using the spike-in of S. pombe transcripts. S. pombe transcripts median reads counts were determined for each sample after the first normalisation step. Then a global mean for S. pombe transcripts reads counts was calculated for quiescent and exponential phase samples. A ratio Exponential/Quiescent was calculated and applied to all Quiescent phase samples (wig files and transcripts reads counts).

Antisens / mRNA coverage was counted and visualized in a -50-+200 nucleotides windows using the Counter RNAseq window (CRAW) package version 0.9.0 (see https://pypi.python.org/pypi/craw/0.9.0 for the program and http://bneron.pages.pasteur.fr/craw/ for the documentation).

Accession number

The data reported here have been deposited in NCBI GEO under the accession number GSE101368.

Acknowledgements

We thank Frank Feuerbach for providing strains LMA2811 and LMA2819.

We thank Cosmin Saveanu and Micheline Fromont-Racine for discussions and critical reading of the manuscript.

This work was supported by the Pasteur Institute (AD, CM, TK and BN), the Centre National de la Recherche Scientifique (AJ and GB). AN received a Fellowship from the French Ministry of Research and the Fondation pour la Recherche Médicale. We acknowledge Jean Yves Coppee and Caroline Proux for the facilities and expertise of the Transcriptomic Platform (PF2) for RNAseq experiments.

Author contributions

AN, AJ and GB designed research. AN, AD and GB performed the experiments. AN, CM, TK, BN AJ and GB analysed the data. AN, AJ and GB interpreted the data and wrote the manuscript.

References and Citations

Alcid EA & Tsukiyama T (2016) Expansion of antisense lncRNA transcriptomes in budding yeast species since the loss of RNAi. Nat. Struct. Mol. Biol. 23: 450–455

Allen C, Büttner S, Aragon AD, Thomas JA, Meirelles O, Jaetao JE, Benn D, Ruby SW, Veenhuis M, Madeo F & Werner-Washburne M (2006) Isolation of quiescent and nonquiescent cells from yeast stationary-phase cultures. J. Cell Biol. 174: 89–100

Aragon AD, Rodriguez AL, Meirelles O, Roy S, Davidson GS, Tapia PH, Allen C, Joe R, Benn D & Werner-Washburne M (2008) Characterization of Differentiated Quiescent and Nonquiescent Cells in Yeast Stationary-Phase Cultures. Mol. Biol. Cell 19: 1271–1280

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

Berretta J, Pinskaya M & Morillon A (2008) A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. Genes Dev. 22: 615–626

Carninci P & Hayashizaki Y (2007) Noncoding RNA transcription beyond annotated genes. Curr. Opin. Genet. Dev. 17: 139–144

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia W-J, Anderson S, Yates J, Washburn MP & Workman JL (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123: 581–592

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

Castelnuovo M & Stutz F (2015) Role of chromatin, environmental changes and single cell heterogeneity in non-coding transcription and gene regulation. Curr. Opin. Cell Biol. 34: 16–22

Castelnuovo M, Zaugg JB, Guffanti E, Maffioletti A, Camblong J, Xu Z, Clauder-Münster S, Steinmetz LM, Luscombe NM & 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

Churchman LS & Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469: 368–373

Criscuolo A & Brisse S (2013) AlienTrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. Genomics 102: 500–506

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW & Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. U. S. A. 103: 5320–5325

van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec 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

Dodt M, Roehr JT, Ahmed R & Dieterich C (2012) FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology 1: 895–905

Donaldson ME & Saville BJ (2012) Natural antisense transcripts in fungi. Mol. Microbiol. 85: 405–417

Erdeniz N, Mortensen UH & Rothstein R (1997) Cloning-free PCR-based allele replacement methods. Genome Res. 7: 1174–1183

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

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H & Oliver SG (1996) Life with 6000 genes. Science 274: 546, 563–567

Gu M, Naiyachit Y, Wood TJ & Millar CB (2015) H2A.Z marks antisense promoters and has positive effects on antisense transcript levels in budding yeast. BMC Genomics 16: 99

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. Mol. Cell 32: 685–695

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

Jensen TH, Jacquier A & Libri D (2013) Dealing with pervasive transcription. Mol. Cell 52: 473–484

Jiang C & Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. 10: 161–172

Keogh M-C, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ, Boone C, Emili A, Weissman JS, Hughes TR, Strahl BD, Grunstein M, Greenblatt JF, Buratowski S & Krogan NJ (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123: 593–605

Kim JH, Lee BB, Oh YM, Zhu C, Steinmetz LM, Lee Y, Kim WK, Lee SB, Buratowski S & Kim T (2016) Modulation of mRNA and lncRNA expression dynamics by the Set2-Rpd3S pathway. Nat. Commun. 7: 13534

Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359

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

Liu Y, Schröder J & Schmidt B (2013) Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. Bioinforma. Oxf. Engl. 29: 308–315

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:

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

McKnight JN, Boerma JW, Breeden LL & Tsukiyama T (2015) Global Promoter Targeting of a Conserved Lysine Deacetylase for Transcriptional Shutoff during Quiescence Entry. Mol. Cell 59: 732–743

Murray SC, Haenni S, Howe FS, Fischl H, Chocian K, Nair A & Mellor J (2015) Sense and antisense transcription are associated with distinct chromatin architectures across genes. Nucleic Acids Res. 43: 7823–7837

Murray SC, Serra Barros A, Brown DA, Dudek P, Ayling J & Mellor J (2012) A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. Nucleic Acids Res. 40: 2432–2444

Nadal-Ribelles M, Solé C, Xu Z, Steinmetz LM, de Nadal E & 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 LM & Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature 457: 1038–1042

Pinskaya M, Gourvennec S & Morillon A (2009) H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. EMBO J. 28: 1697–1707

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

Radonjic M, Andrau J-C, Lijnzaad P, Kemmeren P, Kockelkorn TTJP, van Leenen D, van Berkum NL & Holstege FCP (2005) Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon S. cerevisiae stationary phase exit. Mol. Cell 18: 171–183

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

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

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

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

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

Venkatesh S, Smolle M, Li H, Gogol MM, Saint M, Kumar S, Natarajan K & Workman JL (2012) Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. Nature 489: 452–455

Venters BJ & Pugh BF (2009) How eukaryotic genes are transcribed. Crit. Rev. Biochem. Mol. Biol. 44: 117–141

Verzijlbergen KF, Faber AW, Stulemeijer IJ & van Leeuwen F (2009) Multiple histone modifications in euchromatin promote heterochromatin formation by redundant mechanisms in Saccharomyces cerevisiae. BMC Mol. Biol. 10: 76

van Werven FJ, 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

Wery M, Descrimes M, Vogt N, Dallongeville A-S, Gautheret D & Morillon A (2016) Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. Mol. Cell 61: 379–392

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

Xu Z, Wei W, Gagneur J, Clauder-Münster S, Smolik M, Huber W & Steinmetz LM (2011) Antisense expression increases gene expression variability and locus interdependency. Mol. Syst. Biol. 7: 468

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W & Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. Nature 457: 1033–1037

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

Figures

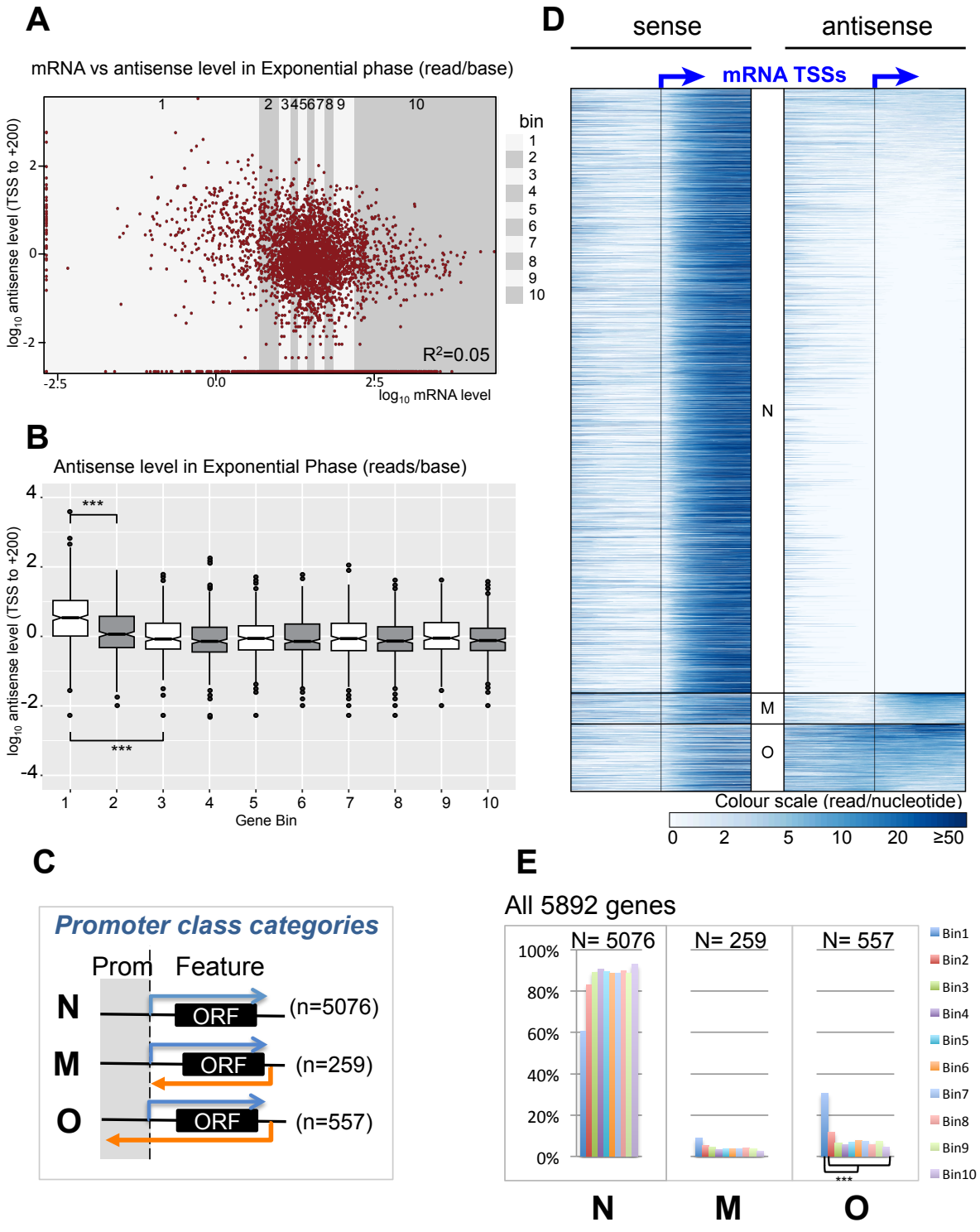


Figure 1 - Antisense ncRNAs are over-represented in lowly expressed genes.

A Scatter plot representing the antisense level (ordinate) function of the corresponding mRNA level (abscissa) in log10 read/base. Gene read count was determined for 5892 genes, and

divided into ten bins (grey strips) of equal length (N = 589 genes per bin for bin1 to bin9; N=591 genes for bin10). The Pearson correlation coefficient R2 = 0.05.

B Comparison of the average antisense level distribution between bins. Boxplots shows the distribution of the average antisense levels within each bin. Brackets indicate the results of an Anova test on pairs of distributions, with *** for $P < 0.001$.

C Schematic of the gene associated promoter class categories depending of the presence and the characteristics of asRNA : N = No asRNA, M = asRNA within the mRNA, O = TSS-overlapping asRNA.

D Heatmap distribution of mRNA and antisense around the TSS of all genes, sorted by antisense and promoter class categories. Depending the class of promoter defined in C, a category N, M or O was assigned to each gene.

E Promoter class categories count per bin. The total number of genes that belong to each class of promoter is indicated (Class N : N=5076; Class M : N=259; Class O : N= 557). Bar charts represent the percentage of each class within the 10 bins defined in A (see also dataset 1). Brackets indicate the results of a statistical inference test on pairs of distributions between bin1 and each other bins, with *** for $P < 0.001$.

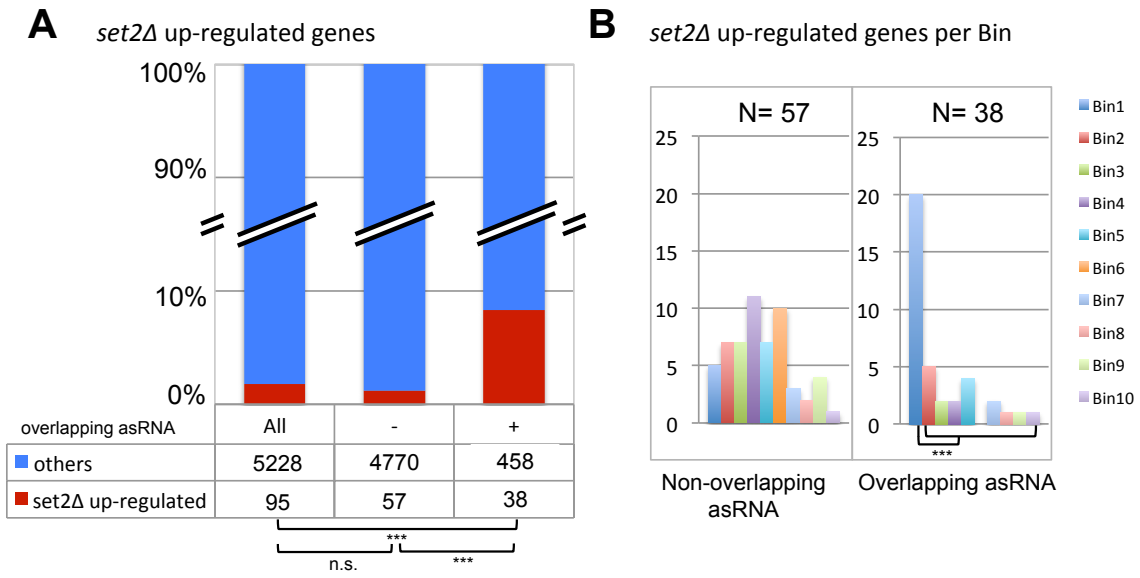


Figure 2 - Promoter overlapping antisenses are overrepresented in set2Δ targets.

A Gene distribution across promoter categories in set2Δ up-regulated genes (dataset from Malabat et al, 2015). Stacked histograms represent the proportion of set2Δ up-regulated genes across all genes (All), or depending the presence of a TSS-overlapping asRNA ("+" = class O) or not ("-" = classes N+M). Brackets indicate the results of a statistical inference test on pairs of distributions, with *** for $P < 0.001$.

B Set2Δ up-regulated genes count depending on the presence of a TSS-overlapping asRNA or not and per bin. Brackets indicate the results of a statistical inference test on pairs of distributions between bin1 and each other bins, with *** for $P < 0.001$

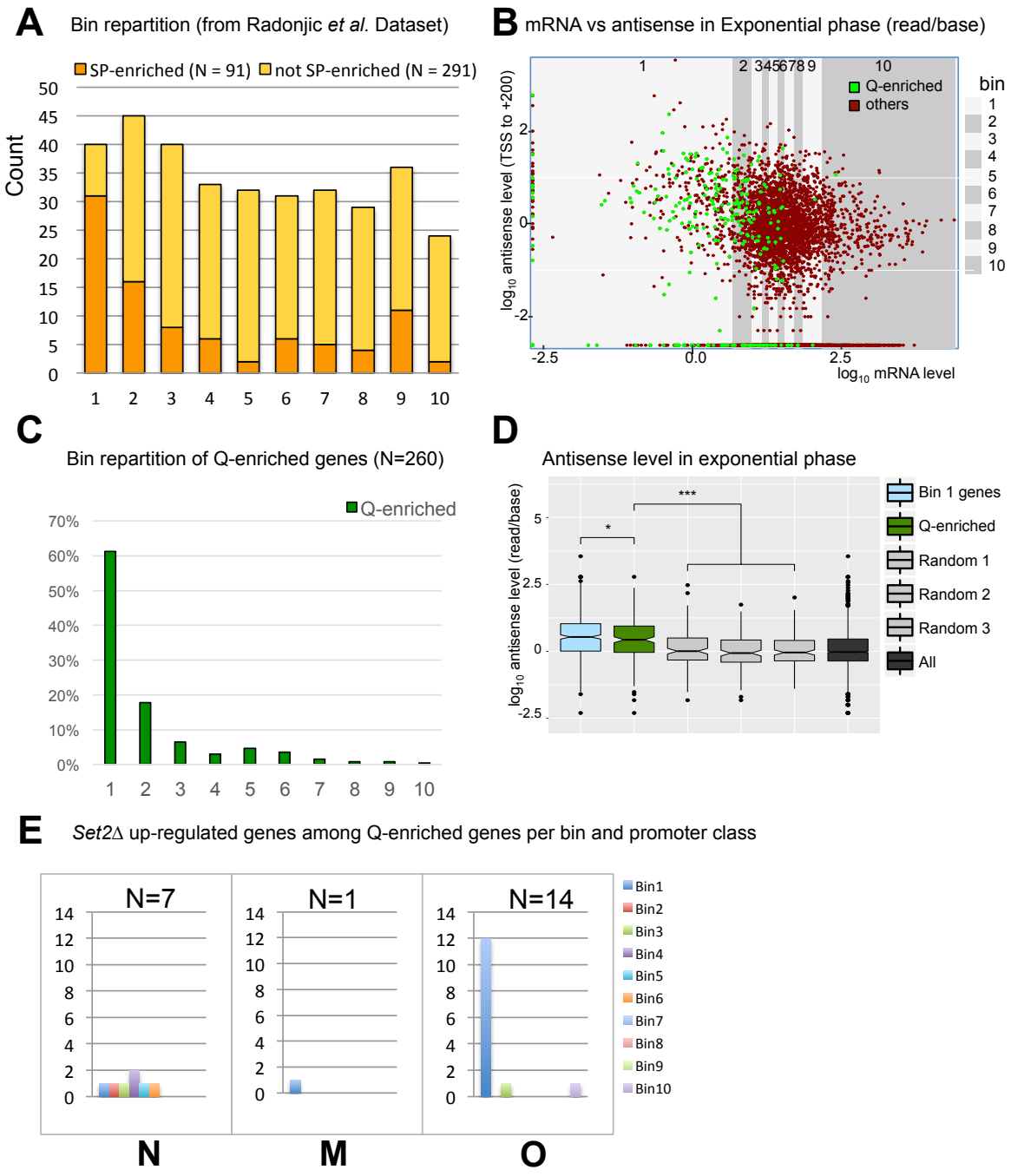


Figure 3 - Quiescent-enriched genes are associated with high antisense level.

A Bar plot of stationary phase-enriched genes count (SP-enriched) versus other genes count (not SP-enriched) within each bin (dataset from Radonjic et al, 2005).

B Distribution of Quiescence-enriched genes among the 5892 yeast genes. Scatter plot of the antisense level as a function of the corresponding mRNA level. 260 genes were found

enriched at least 5 times between exponential and quiescence, defining the Quiescence-enriched genes (Q-enriched, green dots)

C Bar chart of the 260 Q-enriched genes within the 10 bins.

D Distributions of antisense level for different gene categories in exponential phase. Boxplots show the mean antisense level of 260 corresponding Q-enriched genes (green) or "Random" (grey) genes. Random-1, -2, and -3 were defined by random sampling of 260 genes among all the 5892 genes. "Bin_1" (blue) or "All" categories (black) are the measures of all 589 genes from bin1 or all 5892 genes respectively. Brackets indicate the results of an Anova test on pairs, with * for $P < 0.05$, and *** for $P < 0.001$.

E set2Δ up-regulated genes count among Q-enriched genes per promoter class and bin. The bar charts represent the count of set2Δ up-regulated genes within each category of promoter and each bin (see also Dataset 1).

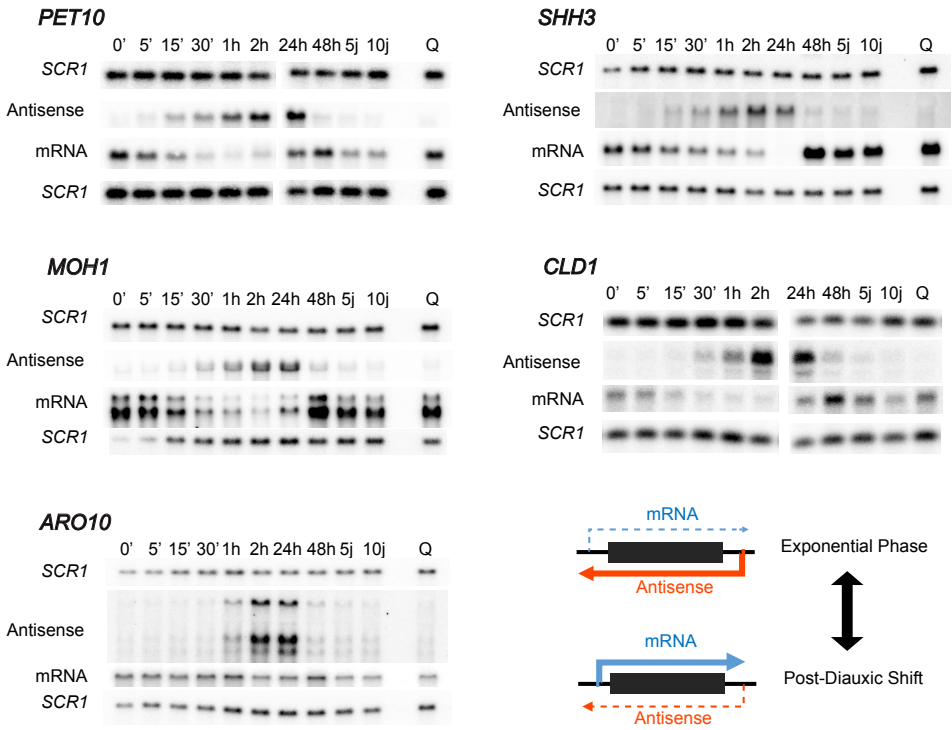


Figure 4 - Q-enriched genes mRNA and corresponding asRNAs are anti-regulated.

Northern-Blot expression time course of mRNA and antisense transcripts in a Δupf1 strain for five examples of Q-enriched genes : PET10, SHH3, MOH1, CLD1 and ARO10. Time point 0' is the time at which quiescent-arrested cells are restarted in rich YPD medium. SCR1 is used as a loading control. RNA probes are described in Figure EV4 and Table EV4.

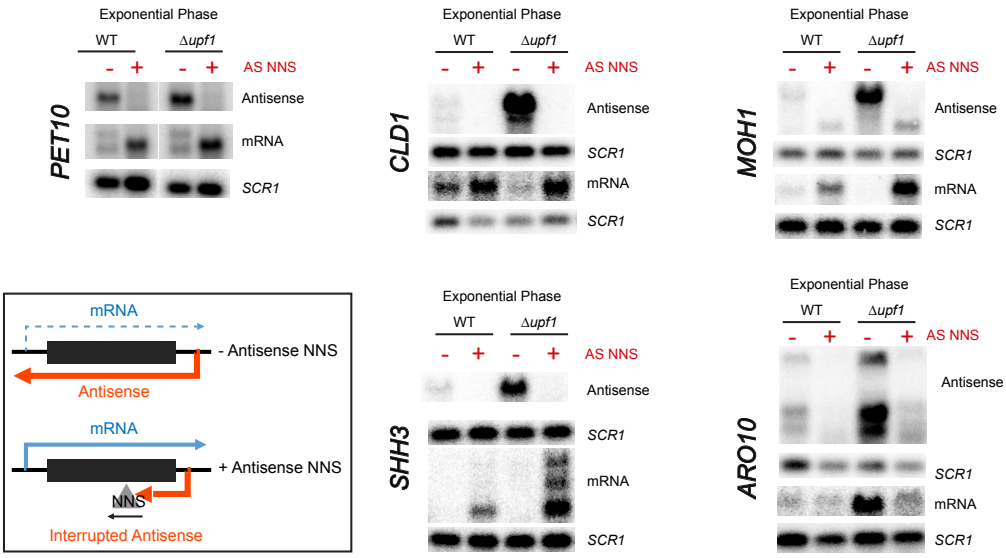


Figure 5 - Antisense transcription interruption during the exponential growth relieves repression of quiescence-enriched genes.

Northern blot analysis of PET10, CLD1, MOH1, SHH3 and ARO10 mRNA and antisense RNAs in the WT and Δupf1 strains with (+) or without (-) the insertion of an antisense Nrd1-Nab3-Sen1 terminator (AS NNS). SCR1 is used as a loading control. RNA probes and NNS insertion are described in Figure EV4 and Table EV4 (see also Material and Methods for strain construction and AS NNS-corresponding strains in Appendix table 1).

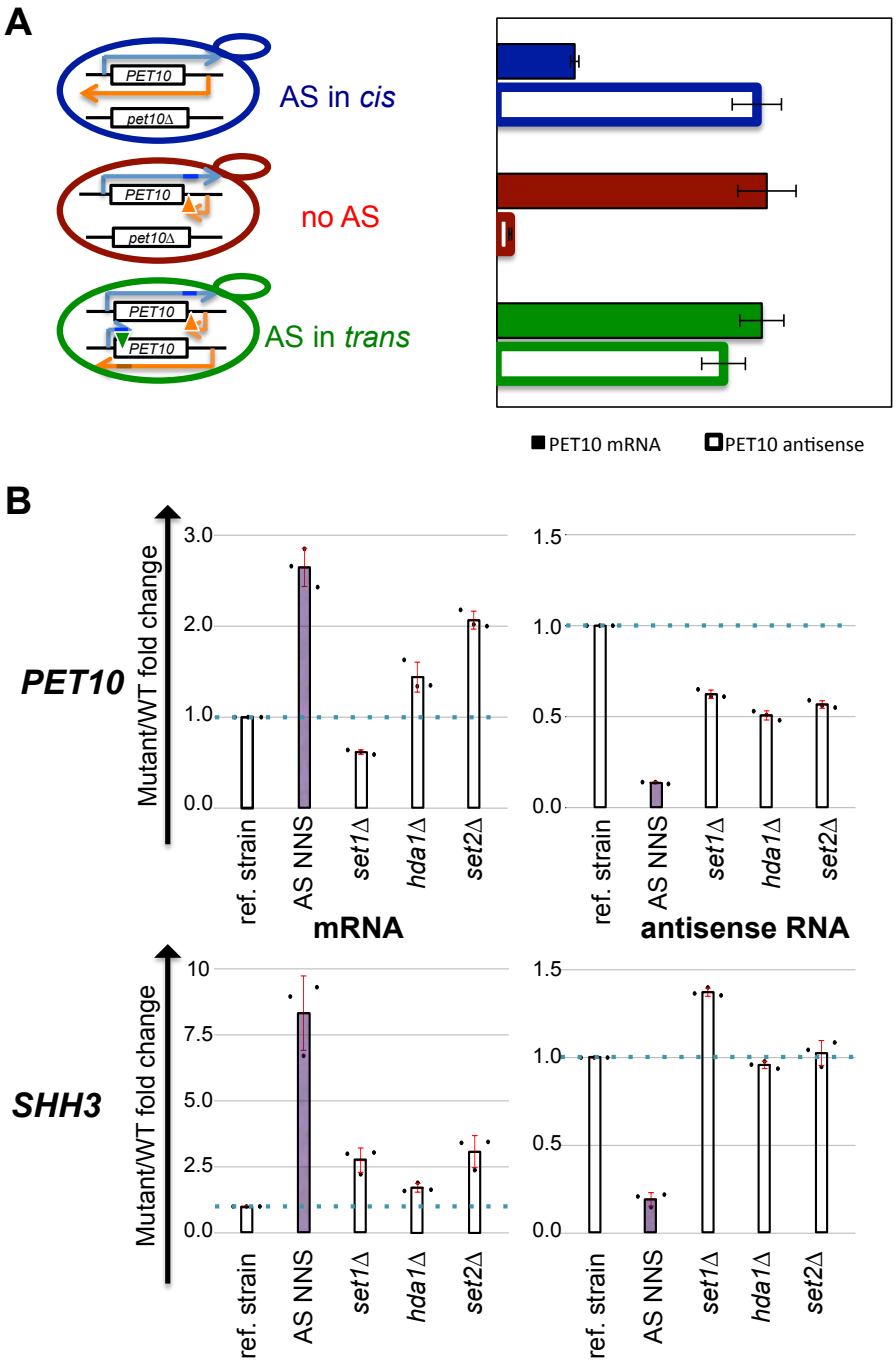


Figure 6 - Antisense repression is mediated by transcriptional interference mechanisms.

A Strand-specific RT-qPCR analysis of PET10 mRNA and antisense RNA abundance in diploid strains. PET10 antisense is transcribed in cis (blue), in trans (green) or not produced (red).

B Strand-specific RT-qPCR analysis analysis of PET10 (upper panel) and SHH3 (lower panel) mRNAs and antisense abundances in a mutant strain where the deletion of UPF1 (ref.

strain) is either combined to an antisense NNS terminator insertion (AS NNS, purple; positive control), or to the deletion of a chromatin modification factor (set1Δ, hda1Δ and set2Δ).

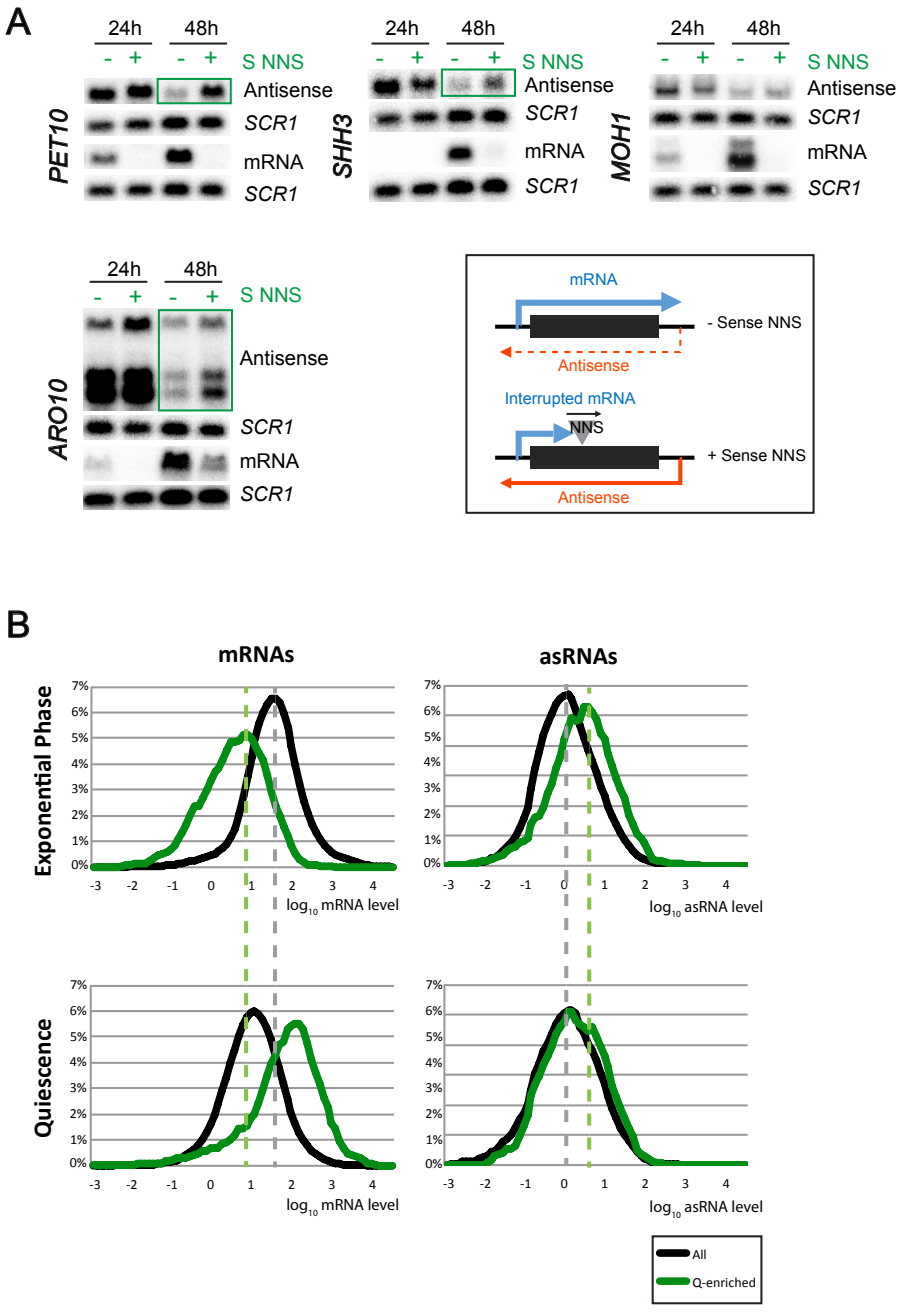


Figure 7 - Gene expression is repressive for antisense non-coding transcription.

A Northern blot analysis of PET10, MOH1, SHH3 and ARO10 mRNA and antisense RNAs in Δupf1 strain, after 24h or 48h of growth in YPD and with (+) or without (-) the insertion of a sense Nrd1-Nab3-Sen1 terminator (NNS S). RNA probes and NNS insertion are described

in Figure EV4 and Table EV4 (see also Material and Methods for strain construction) and S NNS-corresponding strains in Appendix table 1. SCR1 is used as a loading control.

B Comparison of density plots between all (black lines) and Q-enriched genes (green lines) for mRNAs (left panels) or associated asRNA (right panels) from cultures harvested in exponential (upper panels) or G0 (lower panels) phases. Log10 RNA levels are plotted (abscissa) function of the frequency (ordinate).

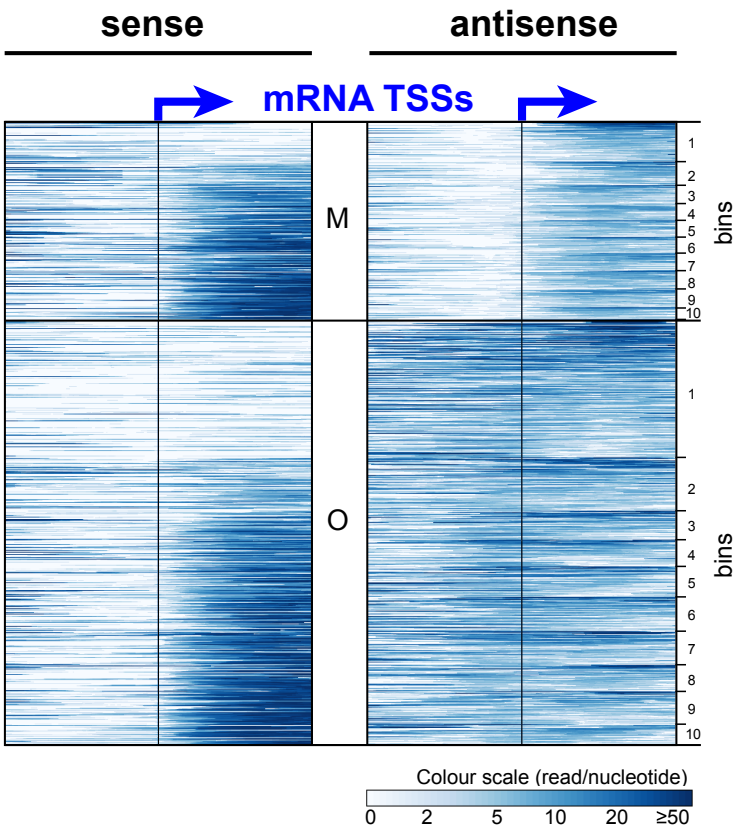
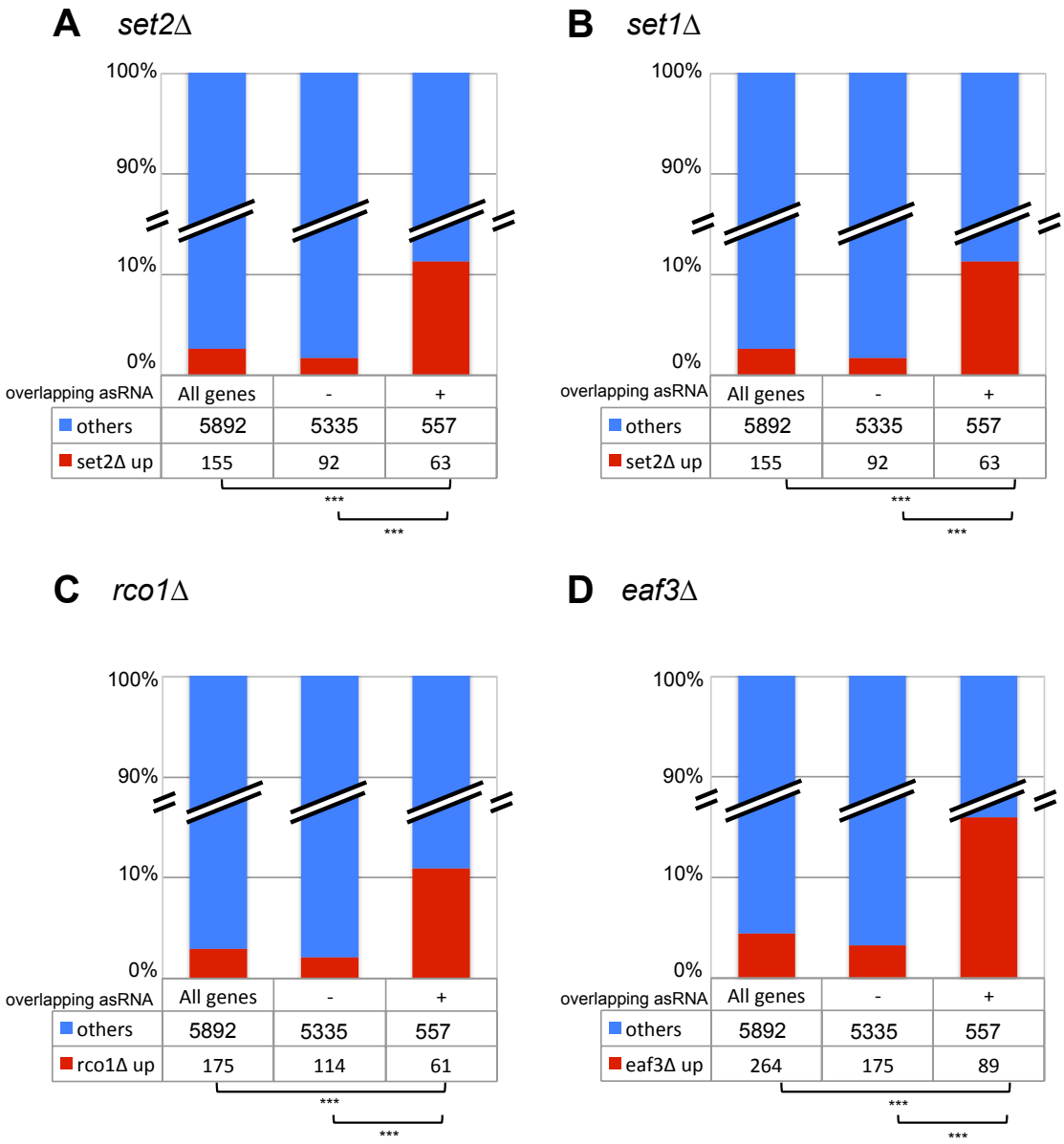


Figure EV1 - Heatmap of asRNA containing genes categories sorted by bin.



E

	set2Δ up	*set1Δ* up	*rco1Δ* up	*eaf3Δ* up
set2Δ up	155			
set1Δ up	80	294		
rco1Δ up	112	86	175	
eaf3Δ up	125	132	139	264

Figure EV2 - Promoter overlapping asRNA associated genes are overrepresented in chromatin modifiers mutants up-regulated targets (datasets from Churchman & Weissman, 2011)

A, B, C, D : Gene count across promoter categories in set2Δ up-regulated genes (dataset from Malabat et al, 2015). Stacked histograms represent the proportion of respectively set2Δ,

set1Δ, rco1Δ and eaf3Δ up-regulated genes across all genes (All genes), or depending the presence of a TSS-overlapping asRNA ("+" ; class O) or not ("-" ; classes N+M). Brackets indicate the results of a statistical inference test on pairs of distributions, with *** for $P < 0.001$

E Overlap of up-regulated genes between the different mutants (respectively set2Δ, set1Δ, rco1Δ and eaf3Δ) in (Churchman & Weissman, 2011 datasets). The table represents the gene overlap between each chromatin modifiers mutants (see also Dataset 1).

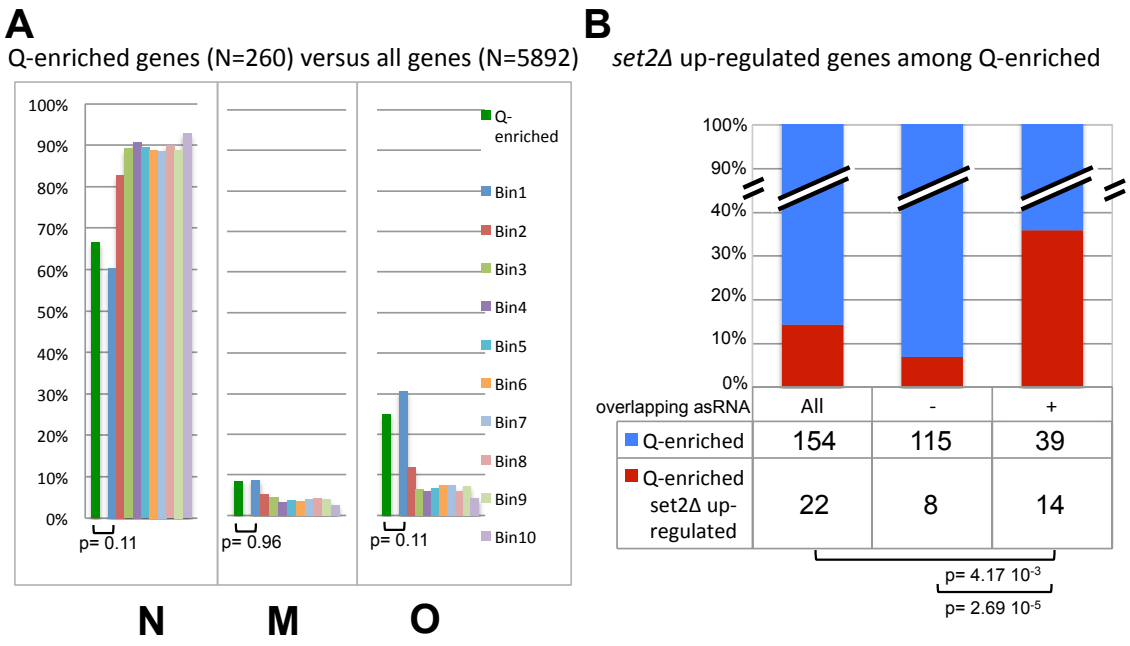


Figure EV3 - Q-enriched genes are more enriched in TSS-overlapping asRNA containing genes.

A Comparison of promoter class categories count per bin between Q-enriched genes and all genes.

The total number of genes that belong to each category is indicated (Q-enriched : N=260, All genes : N=5892). The Bar charts represent the percentage of each class within either the 260 Q-enriched or the 10 bins shown Figure 1E (see also dataset 1). Brackets indicate the results of a statistical inference test on pairs of distributions between bin1 and each other bins, with "n.s." for not significant ($P > 0.05$), ** for $P < 0.01$, and *** for $P < 0.001$.

B Comparison of set2Δ up-regulated genes between Q-enriched genes and all genes Stacked histograms represent the proportion of set2Δ up-regulated genes across all genes (All), or depending the presence of a TSS-overlapping asRNA ("+" = class O) or not ("-" = classes N+M). Brackets indicate the results of a statistical inference test on pairs of distributions, with ** for $P < 0.01$ and *** for $P < 0.001$.

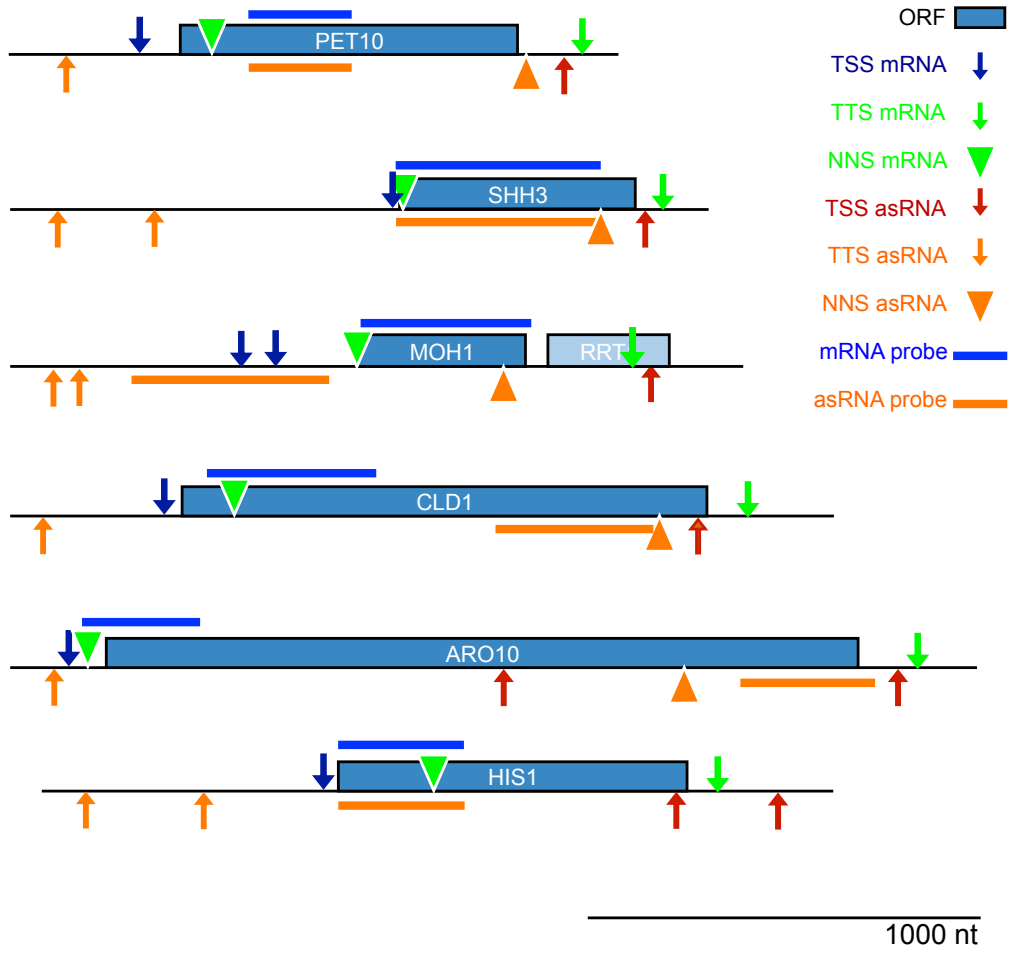


Figure EV4 – Schematic organisation around PET10, SHH3, MOH1, CLD1 and ARO10 loci.

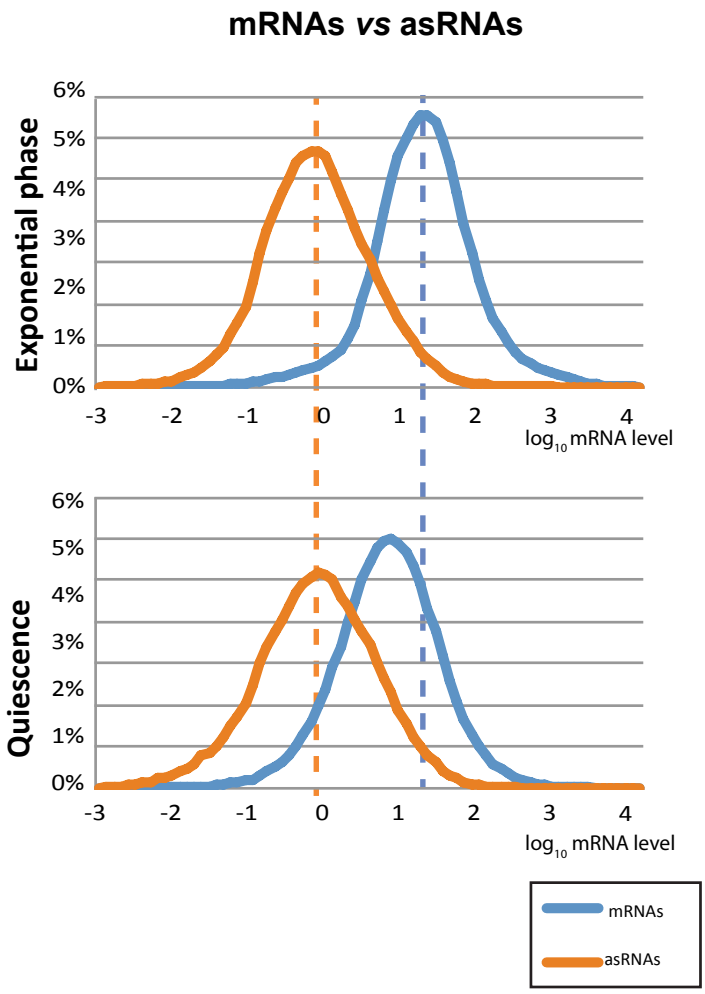


Figure EV5 – mRNA and asRNAs global levels of expression in exponential phase and G0.

Comparison of density plots between for mRNAs (blue lines) or associated asRNA (orange lines) from cultures harvested in exponential (upper panels) or G0 (lower panels) phases. Log10 RNA levels are plotted (abscissa) function of the frequency (ordinate).

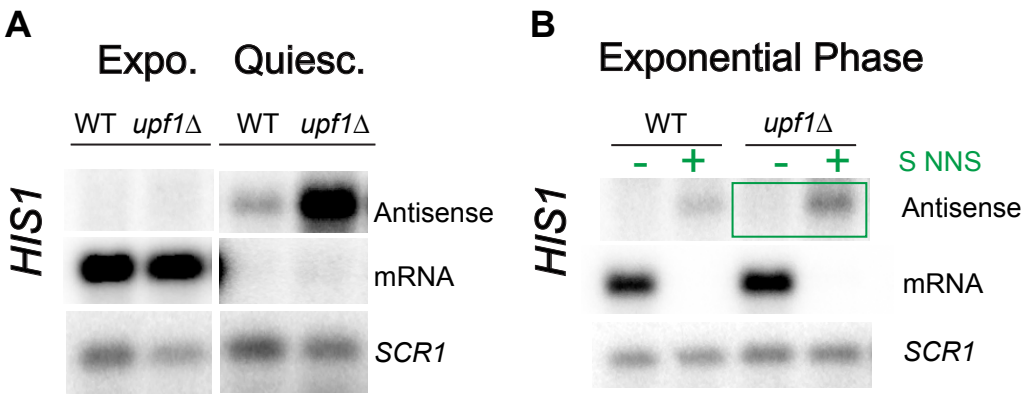


Figure EV6 – HIS1 associated asRNA is induced in quiescence or when HIS1 mRNA transcription is interrupted.

A, B Northern blot analysis of HIS1 mRNA and antisense RNAs in WT and Δupf1 strains in exponential phase or quiescence (A) or in exponential phase with (+) or without (-) the insertion of a sense Nrd1-Nab3-Sen1 terminator (NNS S ; B). RNA probes and NNS insertion are described in Figure EV4 and Table EV4 (see also Material and Methods for strain construction) and S NNS-corresponding strains in Appendix table 1. SCR1 is used as a loading control.

Appendix Supplementary Methods

Yeast strains and cultures

All strains are derivative of BY4741 or BY4742 and were obtained from the Euroscarf deletion collection (http://www.euroscarf.de/). A 37 nucleotides sequence constituting the NNS terminators (GTAATGAATTAAGTCTTGATATATAACAATTAGCTTG construct 78-wt in (Porrua et al, 2012)) was inserted into BY4741 or BY4742 strains using the seamless cloning-free PCR-based allele replacement methods as described in (Erdeniz et al, 1997).
Briefly, gene-specific PCR products containing adaptamer A or adaptamer B and NNS terminator were reconstituted with two successive PCR using A-GENE primer and GENE_NNS_S/AS_rev (PCR1) and GENE_B and NNS_AS_GENE_fwd (PCR2), followed by A_GENE and GENE_B (PCR3). GENE stands for ARO10, PET10 NNS_S, SHH3 MOH1 and CLD1 (see Appendix Table 2). In parallel Fragment L and R were obtained using primer CS1199/CS1200 and CS1201/CS1202 on a URA3 K. lactis DNA template from plasmid pBS1539, (Rigaut et al, 1999). All PCR were done with a high-fidelity Phusion® High-Fidelity (NEBiolabs), following the manufacturer's instructions.
For PET10_NNS antisense, a PCR product GB988/GB989 obtained using pFL38 (from http://seq.yeastgenome.org/vectordb) as a DNA template was used to transform BY4741 plated on SC-URA medium. [URA3+] clones were transformed with 100 pmol of annealed GB990/GB991 primers, plated on YPD at 30°C overnight, and replicated on 5FOA medium in order to select URA3 popped-out constructs. All the constructs were sequence-verified.
Appendix Figure 1 lists the position of NNS terminator insertion, in both sense and antisense orientation. Strains and oligonucleotides are listed in Appendix Table1 and Table EV4 respectively.

Cells were grown to mid-exponential phase in YPD-rich medium at 30 °C, and Quiescent cells were obtained after 10 days of culture at 30°C in YPD and purification on percoll gradient according (Allen et al, 2006).

RNA extraction

Total RNA from logarithmic and G0 cells were extracted with guanidium thiocyanate phenol-chloroform following (Chomczynski & Sacchi, 2006) with the addition of 500 500 µl of glass

beads prior solution D addition, and vortex in a MagNA lyser (Roche) 90 secondes at 4800 rpm after addition of solution D.

Libraries preparation

3' Long SAGE libraries were constructed as described in (Neil & Jacquier, 2011), except than total RNA were extracted from BY4741 logarithmic and G0 cells using the guanidium thiocyanate phenol-chloroform procedure described in (Chomczynski & Sacchi, 2006)

TruSeq stranded mRNA LT sample prep kits (Illumina) were used to prepare RNAseq libraries, on RiboZero gold (Illumina) treated RNA according the manufacturer's instruction. Single read 50 (SR50) sequencing were performed on an Illumina Hiseq 2500 (Pasteur Transcriptomic Platform PF2).

Northern blot

Northern blots were carried out on 4 µg Total RNA as described in (Neil et al, 2009) using strand specific 32P-labeled riboprobes (see Table EV5) except for SCR1 for which a 32P-labeled oligonucleotide was used (GB987 in Table EV5).

Strand-specific RT-qPCR

DNase-treated RNA from exponentially growing yeast cells was used as an input for reverse transcription using 2 pmol of each gene-specific primers and 1 µg RNA using SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions but supplemented with 20 mg/ml actinomycin D to ensure strand specificity of the reverse transcription. For PET10 and SHH3 strand specific reverse transcription, a mix of 2 µM FF3033, AC407, AC429, AC500 and GB1038 primers was used for sense-specific, and FF3033, AC430, AC431 and AC63 primers for antisense-specific measurement (see Table EV5 for gene correspondence). For qPCR, cDNA samples and -RT controls were diluted 10 times, and 2µl were amplified using the qPCR Mix 2X Lo-Rox (Eurobiogreen). CPS1 mRNA was used as the reference gene as its level does not change between exponential and G0 phases.

References

Allen C, Büttner S, Aragon AD, Thomas JA, Meirelles O, Jaetao JE, Benn D, Ruby SW, Veenhuis M, Madeo F & Werner-Washburne M (2006) Isolation of quiescent and nonquiescent cells from yeast stationary-phase cultures. J. Cell Biol. 174 89–100

Chomczynski P & Sacchi N (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. Nat. Protoc. 1 581–585

Erdeniz N, Mortensen UH & Rothstein R (1997) Cloning-free PCR-based allele replacement methods. Genome Res. 7 1174–1183

Neil H & Jacquier A (2011) Enrichment of unstable non-coding RNAs and their genome-wide identification. Methods Mol. Biol. Clifton NJ 759 87–106

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

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

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

Discussions et perspectives

La transcription génique et la transcription pervasive, bien qu'issues du même mécanisme de transcription par l'ARN polymérase II, diffèrent en plusieurs points et notamment par leur origine. En effet, malgré l'existence de certains exemples de transcription pervasive activée par le recrutement de facteurs de transcription spécifiques, il apparaît en fait que la majorité des événements d'initiation de transcription pervasive profitent de l'accès à la chromatine, rendu possible par la transcription génique. Ces transcrits sont ainsi produits à partir de régions pauvres en nucléosomes (promoteurs, terminateurs, ou zones intra-géniques en cours de transcription) (Neil *et al*, 2009; Xu *et al*, 2009; Murray *et al*, 2012; Malabat *et al*, 2015; Venkatesh *et al*, 2016; Kaplan *et al*, 2003), certains étant même produits à partir de GTF activement recrutés pour la transcription génique (Neil *et al*, 2009; Jacquier, 2009). Cette différence d'initiation transcriptionnelle entre la transcription codante et non-codante se traduit par un niveau globalement faible de la transcription pervasive, apparaissant comme une transcription basale envahissante et bruitée.

Ces deux types de transcrits s'opposent également par leur devenir dans la cellule. Le but de la transcription génique, activée de façon constitutive ou en réponse à un stimulus extérieur, est d'être productive et finement régulée, de façon à garantir la disponibilité de ses produits protéiques fonctionnels quand ils sont requis. La transcription pervasive est pour sa part ciblée par des mécanismes de contrôles développés par la cellule, dès le stade de l'initiation et jusqu'au cytoplasme, pour conduire à la dégradation des transcrits qui en sont issus. En particulier, cette transcription non-codante est ciblée par un contrôle nucléaire provoquant une terminaison précoce liée à une dégradation des transcrits, complété par un deuxième contrôle cytoplasmique, utilisant la voie du NMD (Thiebaut *et al*, 2006; Gudipati *et al*, 2008; Porrua *et al*, 2012; Porrua & Libri, 2015; Malabat *et al*, 2015; Wery *et al*, 2016).

Malgré ces différences d'origine et de destin cellulaire, les transcriptions codantes et non-codantes utilisent le même mécanisme de transcription pour leur élongation. Elles doivent de la même manière accéder à la matrice d'ADN compactée dans les nucléosomes et laissent la même signature sur la chromatine. Pour les deux types de transcription, des facteurs de modifications de la chromatine sont effectivement recrutés co-transcriptionnellement, par exemple en fonction des modifications du CTD de la polymérase. Au vu de l'absence de

capacité codante suffisante, et du ciblage presque automatique de la transcription pervasive par la machinerie de dégradation cellulaire des ARN, la question de sa fonctionnalité globale n'est pas complètement résolue. Cependant, l'existence de contrôles cellulaires pour restreindre cette transcription suggère qu'elle possède le potentiel de perturber l'expression de gènes au niveau transcriptionnel, ce qui a récemment été démontré au niveau global (Schulz *et al*, 2013).

Chez *S. cerevisiae*, les exemples ponctuels de fonctionnalité de la transcription pervasive ont principalement mis en évidence un rôle en *cis* de la transcription non-codante lié aux modifications de la chromatine engendrées (Partie III de l'Introduction). En particulier, on distingue deux types d'effets des modifications de la chromatine dus à la transcription pervasive antisens. Le premier est global, et intervient principalement pour réguler la cinétique d'induction ou pour augmenter le bruit transcriptionnel des gènes associés à un transcrits antisens (Kim *et al*, 2012; Xu *et al*, 2011; Murray *et al*, 2015). Le deuxième effet est spécifique à certains gènes et a été décrit pour des exemples ponctuels. Il intervient majoritairement pour réprimer l'expression des gènes *via* un mécanisme d'interférence transcriptionnelle (Tableau 1).

Mon projet d'étude de l'impact des modifications de la chromatine liées à la transcription antisens non-codante (en Master 2) va également dans le sens de l'existence de deux niveaux de régulation par la transcription antisens. En effet, en utilisant un système d'induction de multiples phénomènes d'interférence transcriptionnelle (dans la souche tetNrd1*upf1*Δ), j'ai analysé les conséquences de la délétion de facteurs de modifications de la chromatine sur la répression des gènes. J'ai pu mettre en évidence que ces facteurs étaient impliqués soit pour des régulations très faibles mais globales de l'expression des gènes, soit pour des régulations plus fortes spécifiques au gène régulé (Figure 15, Partie I.1 des Résultats). Pour environ 30% des 76 gènes fortement régulés par un phénomène d'interférence transcriptionnelle, nous avons pu impliquer les facteurs Set1, Chd1, Hda1 ou Set3 dans le mécanisme de répression. De façon intéressante, nous avons pu remarquer une spécificité de facteur impliqué selon le gène régulé, avec très peu de recouvrement entre les classes (Figure 16). Cela suggère que différentes voies de modifications de la chromatine redondantes interviennent dans le mécanisme d'interférence transcriptionnelle. Cette observation est en accord avec les exemples ponctuels de régulation génique par un antisens non-codant qui impliquaient des voies de modifications de la chromatine diverses (Tableau 1).

L'absence d'implication des facteurs pré-sélectionnés dans la régulation de 70% des gènes restants, nous a amenés à développer un système rapporteur associé à des combinaisons différentes de promoteurs sens et antisens, afin d'identifier d'autres facteurs impliqués en *trans* dans le mécanisme d'interférence transcriptionnelle, *via* un crible génétique. Nous avons malheureusement été confrontés à l'impossibilité de reproduire le phénomène au niveau des systèmes rapporteurs CUP1 et URA3-GFP-mCherry (Figures 20 et 23). Ces problèmes techniques traduisent selon moi la difficulté d'étudier des régulations transcriptionnelles au niveau protéique, et surtout l'éventualité que les propriétés intrinsèques aux régions promotrices sens et antisens ne soient pas suffisantes pour expliquer la complexité du mécanisme d'interférence transcriptionnelle. A ce jour, aucun système rapporteur n'a été développé pour rechercher de manière systématique les facteurs impliqués en *trans* dans le phénomène d'interférence transcriptionelle. En revanche, un système rapporteur a été élaboré par le laboratoire de Stephen Buratowski, pour étudier les facteurs influant sur la bidirectionnalité du promoteur génique, et donc la production d'ARN non-codant divergent. Dans ce système rapporteur, les régions transcrites de l'ARNm et de l'antisens ont été remplacées par la séquence codant pour la YFP pour « Yellow fluorescent protein » ou la mCherry respectivement (Marquardt *et al*, 2014). L'environnement du promoteur génique, potentiellement impliqué dans sa bidirectionnalité, a donc ainsi été conservé.

Le suivi des transcriptions sens et antisens en cellules individualisées pour l'étude de phénomènes d'interférence transcriptionnelle ne nécessite pas forcément de système rapporteur. En effet, l'analyse consiste généralement en une détection des transcrits au niveau de leur site de synthèse et non pas en un suivi des protéines traduites. Diverses études ont été menées dans ce sens en utilisant la technique de FISH pour « Fluorescent In Situ Hybridation » avec des sondes ARN complémentaires marquées par des fluorochromes. Des expériences de FISH en cellules individualisées ont notamment permis de conclure à l'expression bimodale du gène PHO84 et de son antisens dans une souche *rrp6*Δ. Leur transcription est en effet mutuellement exclusive, la transcription de PHO84 étant transitoire ou « en burst », tandis que celle de son antisens est constante et faible (Castelnuovo *et al*, 2013). La même méthode de FISH a permis d'étudier en cellule individuelle la fréquence des transcriptions codantes et non-codantes pour décrire la régulation des gènes *HMS2*, *FLO11*, et *IME1* (Nguyen *et al*, 2014; Bumgarner *et al*, 2012; van Werven *et al*, 2012). Le système « MS2 / PP7 » utilisant la reconnaissance d'une structure secondaire intégrée dans l'ARN

d'intérêt par des protéines (MS2 ou PP7) couplées à des fluorochromes, a également été développé (Bertrand *et al*, 1998). C'est ainsi que la régulation du gène GAL10 par son antisens a été décrite dans une étude de 2015 (Lenstra *et al*, 2015). Ces deux techniques pour la visualisation des ARN dans la cellule sont d'ailleurs avantageuses car elles permettent la détection des transcrits sur leur lieu de synthèse, avant une dégradation éventuelle.

La recherche de cas naturels de régulation de l'expression génique par la transcription pervasive nous a amenés à réaliser une étude comparative de la transcription génique et antisens non-codante, dans un mutant *upf1*Δ, pour invalider la dégradation cytoplasmique des transcrits. L'analyse de résultats de séquençage nous a permis de constater l'absence de corrélation générale entre la transcription sens et la transcription antisens dans un état d'équilibre, en accord avec une analyse précédente réalisée à partir de résultats de Net-Seq (Murray *et al*, 2015). Cependant, il est apparu de nos résultats de RNA-Seq qu'une fraction importante des gènes faiblement exprimés est associée à un fort signal antisens. De plus, ces antisens associés à des gènes faiblement exprimés sont enrichis pour des antisens qui vont au-delà du TSS génique, et qui ont donc un potentiel régulateur plus fort (Figure 1 de l'article). Ces premières observations avaient été formulées par le laboratoire de Lars Steinmetz en 2011, après des expériences de puces à ADNc à haute densité (« Tiling arrays ») réalisées dans une souche sauvage (Xu *et al*, 2011). L'analyse des SUT antisens a effectivement montré que les gènes associés à un signal antisens se caractérisent par un niveau minimal d'expression, qui est inférieur à celui des gènes pour lesquels aucun antisens n'est détecté. Les auteurs ont alors suggéré un rôle de la transcription antisens pour la répression de gènes peu exprimés, répression qui serait levée lors de l'induction génique. De plus, il a également été remarqué que les antisens recouvrant le TSS génique sont associés à des gènes qui sont d'avantage régulés (Xu *et al*, 2011). Par ailleurs, le laboratoire de Françoise Stutz a étudié les antisens accumulés dans une souche *rrp6*Δ, en définissant des gènes dont l'expression est diminuée suite à l'augmentation de transcription antisens en réponse à la délétion de *RRP6*. La définition d'une classe de gènes qui semblent réprimés dans le mutant *rrp6*Δ de façon dépendante de modifications de la chromatine (probablement par interférence transcriptionnelle), a permis d'observer que les antisens de ces gènes ont tendance à aller au-delà du TSS du gène (Castelnuovo *et al*, 2014).

Le fait qu'il y ait plus d'antisens lorsque l'expression génique est plus faible peut potentiellement refléter leur fonction répressive. Mais cela peut aussi être en partie expliqué

par l'absence de répression de l'initiation de la transcription pervasive par la transcription génique si celle-ci est trop faible (Introduction partie II.1). En faveur de la première hypothèse, l'analyse de résultats d'études transcriptomiques publiées (Churchman & Weissman, 2011; Malabat *et al*, 2015) a montré que les gènes significativement augmentés en l'absence des facteurs de modification de la chromatine Set2, Set1, Rco1, et Eaf3 sont également enrichis en antisens qui recouvrent le TSS génique (Figures 2 et EV2 de l'article). Cela suggère qu'une partie au moins de ces gènes pourrait être réprimée par un mécanisme d'interférence transcriptionnelle lié à la transcription antisens.

Afin d'identifier un rôle de la transcription pervasive antisens plus général que le rôle jusqu'alors décrit sur des exemples spécifiques, nous avons recherché l'existence de phénomènes d'interférence transcriptionnelle multiples, pour la répression d'un ensemble de gènes co-régulés entre deux conditions. En effet, nous avons voulu identifier un mécanisme global de régulation en réponse à une variation environnementale, et non des cas ponctuels et indépendants de régulation génique comme décrits précédemment (Tableau 1).
Nous avons pu remarquer que les gènes caractérisés par une faible expression en phase exponentielle et par une association à de nombreux antisens recouvrant leur TSS, comprenaient une grande partie des gènes dont les ARN sont enrichis en phase stationnaire (« SP-enriched ») décrits dans une étude précédente (Radonjic *et al*, 2005). Nous avons choisi d'étudier le contexte de la quiescence contenant une population cellulaire plus homogène que la phase stationnaire, et permettant de visualiser la résultante des régulations géniques liées à un appauvrissement naturel en nutriments de l'environnement de *S. cerevisiae*. La comparaison de résultats de séquençage entre deux états d'équilibre, en phase exponentielle ou en quiescence, nous a ainsi permis de définir des gènes enrichis en quiescence. Ces gènes regroupent toutes les caractéristiques des gènes faiblement exprimés en phase exponentielle, à savoir qu'ils sont associés à un fort niveau d'antisens potentiellement régulateurs (Figure 3 de l'article). De plus, l'induction de ces gènes en quiescence est associée à une diminution des antisens associés (Figure 4 de l'article).

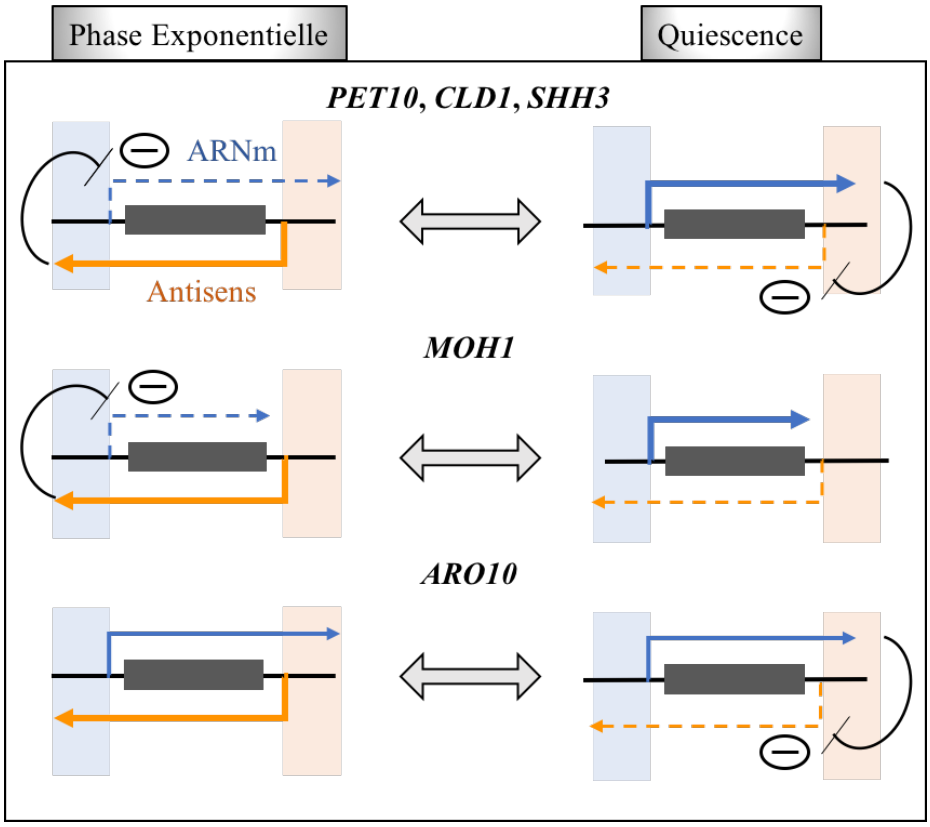


Figure 24 : Récapitulatif des régulations déduites des expériences d'interruption NNS pour PET10, CLD1, SHH3, MOH1, et ARO10 *(Les rectangles bleus et oranges correspondent respectivement aux promoteurs géniques et antisens)*

Des expériences d'interruption directionnelle de la transcription ont été réalisées sur des exemples de gènes dont les ARN sont enrichis en quiescence : *PET10*, *MOH1*, *SHH3*, *CLD1* et *ARO10*. En utilisant le mode de terminaison naturel des CUT (Porrua *et al*, 2012) nous avons pu confirmer l'existence d'une régulation mutuelle des transcriptions sens et antisens (Figures 5 et 7 de l'article) précédemment proposée (Xu *et al*, 2011). En effet, lorsque les gènes sont très faiblement exprimés en phase exponentielle, la transcription antisens recouvrant leur promoteur conduit à leur répression en *cis*. Lorsque les conditions environnementales conduisent à l'induction des gènes, l'activation génique permet à la transcription génique de surpasser la répression par la transcription antisens et de la réprimer à son tour. Cela est conditionné par le fait que la transcription sens aille au-delà du TSS du transcrit antisens. Ce modèle est en accord avec les régulations sens/antisens des gènes *PET10*, *CLD1* et *SHH3* entre la phase exponentielle et la quiescence (Figure 24, panel du haut). *ARO10* est le seul des cinq gènes que nous avons étudié pour lequel nous n'avons pas pu montrer de rôle répresseur de la transcription antisens. Mais les cinétiques de transcription sens et antisens montrent que son ARNm n'est pas fortement réprimé en phase exponentielle dans nos conditions de culture (Figure 4 de l'article et Figure 24, panel du bas). Cela est donc

cohérent avec l'absence de rôle répressif de la transcription antisens dans ce contexte. Cependant, il est intéressant de remarquer que *ARO10* est aussi le seul des cinq gènes que nous avons analysé dont la fonction de la transcription antisens a été testée dans analyse à grande échelle indépendante (Huber *et al*, 2016), mais en utilisant des conditions de culture différentes. Or, de manière tout à fait remarquable, cette étude a montré un rôle régulateur de la transcription antisens de *ARO10* pendant la croissance en milieu synthétique complet. Il s'avère donc que, dans les cinq exemples que nous avons analysés, on peut mettre en évidence un rôle régulateur pour la transcription antisens.
Enfin, *MOH1* est le seul gène dont la transcription ne réprime pas la transcription antisens, mais c'est aussi le seul des cinq gènes étudiés dont la transcription de l'ARNm ne va pas au-delà du TSS antisens (Figure 24, panel du milieu).

Notre étude a montré que la répression génique par la transcription antisens se produisait en *cis* et impliquait des modifications de la chromatine (Figure 6 de l'article). Une preuve plus directe de la régulation par interférence transcriptionnelle pourrait être apportée par une étude des marques de modifications de la chromatine spécifiquement déposées par la transcription antisens au niveau des promoteurs régulés (par des analyses comparatives de CHIP-seq entre les souches sauvages, et NNS antisens par exemple).

Nous avons donc démontré pour la première fois que le phénomène d'interférence transcriptionnelle liée à la transcription antisens représentait un mécanisme de répression d'une large fraction de gènes partageant un même profil de régulation. Ce rôle global permettrait de réprimer la transcription basale de gènes dans des conditions où ils ne sont pas induits. Ces observations sont en accord avec l'idée que la transcription antisens permet d'éviter une fuite transcriptionnelle au niveau de gènes supposés être éteints, comme cela avait été précédemment suggéré (Castelnuovo *et al*, 2013; Lenstra *et al*, 2015).

Il apparaît qu'un certain nombre des gènes bien exprimés sont néanmoins associés à la présence d'antisens chevauchant le promoteur. En effet si les gènes de la classe O sont plus représentés parmi les gènes les moins exprimés (bin 1 dans la figure 1 de l'article), ils restent représentés de manière moindre mais équivalente dans les gènes plus fortement exprimés (bin 2 à 10 ; Figures 1 et EV1 de l'article). Certains gènes gardent donc un niveau d'expression élevé malgré la présence d'un antisens chevauchant. Cela n'exclue pas un potentiel régulateur des antisens associés aux gènes fortement exprimés. En effet, il a été précédemment montré

que la transcription non-codante pouvait jouer un rôle positif pour la transcription génique (Uhler *et al*, 2007; Nadal-Ribelles *et al*, 2014).
En outre, une étude a également décrit que la transcription pervasive pouvait, en fonction des conditions environnementales, avoir un rôle régulateur ou ne représenter que du bruit transcriptionnel (Lenstra *et al*, 2015). La transcription antisens chevauchant le promoteur d'un gène peut donc être répressive dans une condition donnée, et non fonctionnelle dans une autre. C'est ce que nous avons observé avec l'exemple d'ARO10, qui est insensible à la présence d'un antisens dans nos conditions de cultures (YPD) et dont l'antisens est régulateur en milieu synthétique complet (Huber *et al*, 2016). Cette insensibilité condition-spécifique des gènes à l'interférence transcriptionnelle pourrait donc expliquer pourquoi certains gènes non réprimés ont tout de même de la transcription antisens au niveau de leur promoteur.

La transcription antisens et les modifications de la chromatine qu'elle entraîne constituent donc un niveau de complexité supplémentaire dans la régulation de l'expression des gènes. Cette régulation au niveau transcriptionnelle présente l'avantage d'être mise en place plus rapidement qu'une régulation par des facteurs de transcription (qui doivent être exprimés et localisés au bon endroit), et de pouvoir être finement contrôlée grâce la régulation mutuelle seuil-dépendante des transcriptions géniques et antisens. Cependant les deux niveaux de régulation des gènes, *via* les facteurs de transcription et *via* les phénomènes d'interférence transcriptionnelle, sont très probablement interdépendants. Afin de comprendre la transition du profil d'expression sens/antisens entre la phase exponentielle et la quiescence, il serait donc utile de prendre en compte la régulation par des activateurs transcriptionnels qui pourraient être responsables d'une activation extérieure de la transcription génique, après un épuisement en nutriments dans le milieu de culture cellulaire. Cette induction augmenterait ainsi la transcription sens au-dessus de son seuil de sensibilité à l'interférence par la transcription antisens, et lui permettrait de réprimer à son tour la transcription antisens. Il pourrait donc être intéressant d'étudier plus particulièrement les facteurs de transcription qui pourraient être impliqués dans l'induction des gènes *PET10*, *CLD1* et *SHH3*, afin de vérifier cette hypothèse.

Le gène *HIS1*, que nous avons également étudié, présente un comportement inverse à celui des gènes enrichis en quiescence puisqu'il est exprimé spécifiquement en phase exponentielle. Nous avons pu montrer par les mêmes expériences d'interruption directionnelle de la transcription que ce gène est également très légèrement régulé par la transcription antisens

durant la phase de quiescence, lorsque son expression doit être éteinte (Résultat non montré). Le maintien d'un certain niveau de transcription pervasive, malgré l'extinction globale de l'expression génique en quiescence (Figure EV5 de l'article), pourrait donc en partie être expliqué par la fonction répressive de la transcription génique, bien que l'absence de cette dernière, et de son effet répressif sur la transcription antisens, puisse aussi expliquer une émergence accrue de la transcription pervasive à l'échelle génomique.

En résumé, cette étude ouvre la possibilité de l'existence de phénomènes multiples de régulations de l'expression génique par la transcription antisens. Il existe très vraisemblablement d'autres phénomènes de co-régulation de différents gènes par interférence transcriptionnelle, dans d'autres transitions entre différentes conditions environnementales. Nous avons donc pu ici apporter une preuve supplémentaire de la fonctionnalité d'une fraction de la transcription pervasive chez *Saccharomyces cerevisiae.*

Les gènes *PET10* et *MOH1*, leur antisens, et l'anti-régulation entre ces deux transcriptions de la phase exponentielle à la phase stationnaire avaient précédemment été étudiés par l'analyse de leur conservation entre différentes espèces de levures : *Saccharomyces cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces mikatae*, *Saccharomyces bayanus*, *Naumovozyma castellii* et *Kluyveromyces lactis*. Le gène *MOH1* et son antisens ne sont pas retrouvés dans trois espèces et l'anti-régulation sens/antisens ne semble pas conservée. En revanche *PET10*, son antisens et leur anti-régulation entre la phase logarithmique et la phase stationnaire sont très bien conservés, ce qui suggère que la régulation par interférence transcriptionnelle l'est aussi (Yassour *et al*, 2010).

L'étude de la conservation des lncARN est utile pour renforcer ce modèle de fonctionnalité de la transcription non-codante. Ce type d'analyse a été réalisé en 2016 par Jessica Vera et Robin Dowell, en se limitant aux CUT qui ont un potentiel régulateur moindre. Cependant, ces CUT peuvent, selon leur capacité à échapper à la terminaison NNS, être à l'origine de la production de lncARN et acquérir un potentiel régulateur plus important (Castelnuovo *et al*, 2014). La recherche de leur conservation n'est donc pas sans intérêt dans le cadre de l'étude de la fonctionnalité de la transcription non-codante. Dans cette étude, la comparaison des CUT détectés entre plusieurs souches de *Saccharomyces cerevisiae* issues de variétés différentes, et avec une souche de *Saccharomyces paradoxus*, a montré une forte conservation des CUT identifiés (Vera & Dowell, 2016).

Enfin, une autre analyse phylogénétique plus spécifique a été réalisée, selon l'association des gènes à des antisens non-codants. Cette étude d'indicateurs de la complexité de régulation génique a notamment montré que les transcrits antisens associés à des gènes de régulation plus complexe étaient soumis à une plus forte pression évolutive, et qu'ils avaient tendance à présenter un profil d'expression inversé par rapport à celui du gène (Lin *et al*, 2013). Cela est donc cohérent avec l'idée qu'une partie au moins des lncARN antisens a acquis une fonction régulatrice au cours de l'évolution.

Une étude de l'évolution des lncARN antisens a été récemment publiée. La comparaison des résultats de séquençage des ARN chez plusieurs espèces de levures bourgeonnantes (*Naumovozyma castellii*, *Saccharomyces uvarum*, *Saccharomyces kudriavzevii*, *Saccharomyces mikatae* et *Saccharomyces cerevisiae*), a pu montrer que depuis la divergence des *Saccharomyces* à partir de *N. castellii*, le niveau de lncARN antisens a globalement augmenté (Figure 25).

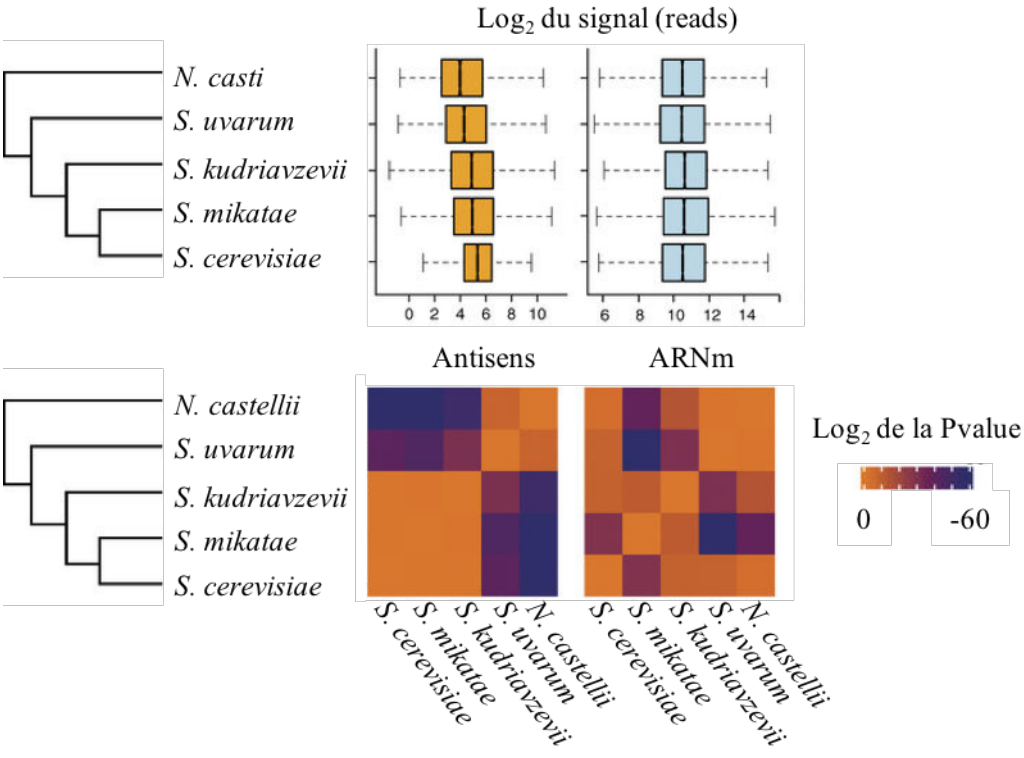


Figure 25 : Représentation de l'évolution du niveau d'Antisens / d'ARNm et des P values associées (Wilcoxon apparié) (Adapté de Alcid & Tsukiyama, 2016)

La taille des lncARN antisens ainsi que la taille du recouvrement du gène associé (donc leur potentiel régulateur) ont également augmenté au cours de l'évolution (Figure 26).

Cette évolution serait expliquée par la diminution de la contrainte évolutive exercée sur les lncARN, à cause de la perte progressive de la machinerie du RNAi. Il semblerait donc que depuis cette perte, les levures aient développé un système alternatif de régulation de l'expression génique à partir d'une transcription pervasive qui ne représentait à l'origine qu'un bruit transcriptionnel contre-sélectionné (Alcid & Tsukiyama, 2016).

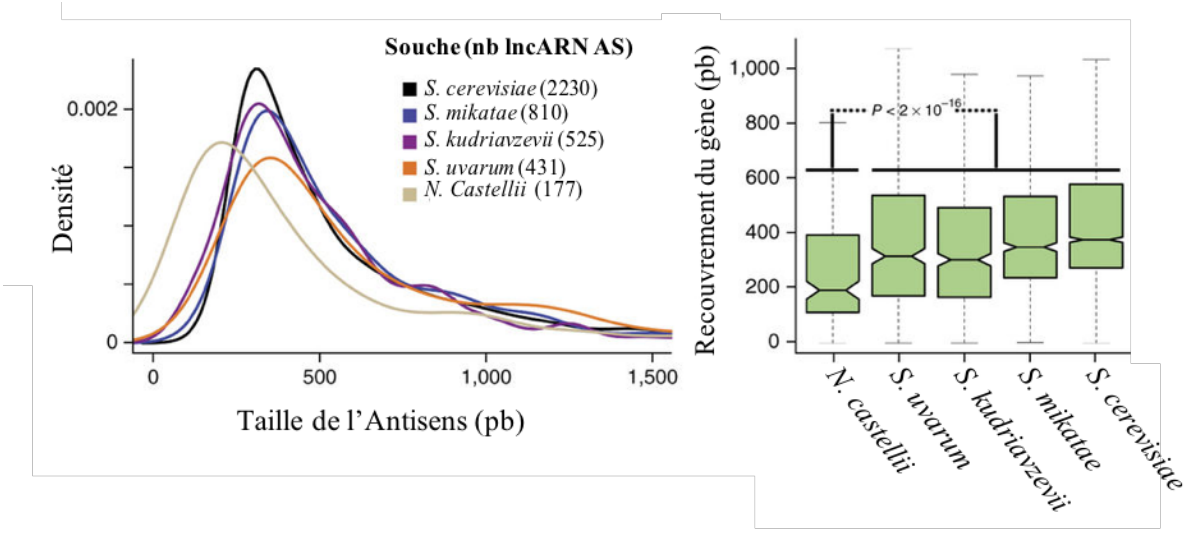


Figure 26 : Evolution de la taille des lncARN et de leur taille de recouvrement avec les gènes associés (Adapté de Alcid & Tsukiyama, 2016)

Conclusion

La question de la fonctionnalité de la transcription pervasive est ouverte depuis sa mise en évidence il y a une dizaine d'années. Les ARN pervasifs correspondent effectivement à une transcription non-codante envahissante, profitant de l'accessibilité de la chromatine permise par la transcription génique. De plus, ces transcrits sont ciblés dans la cellule par des voies de dégradation nucléaires et cytoplasmiques qui suivent leur synthèse.

Cependant, la levure *Sacharomyces cerevisiae,* qui a perdu la machinerie d'ARN interférence, a développé un système de régulation génique au niveau transcriptionnel impliquant cette transcription pervasive. Certains lncARN ont ainsi acquis au cours de l'évolution un rôle répresseur des gènes *via* un phénomène d'interférence transcriptionnelle. Ce rôle répresseur se traduit par une inhibition de l'initiation de la transcription génique par les marques de modifications de la chromatine déposées par la transcription non-codante. Afin d'éviter que la transcription pervasive n'interfère systématiquement avec l'expression génique, la cellule a développé un système de terminaison précoce de la transcription non-codante utilisant la voie de terminaison Nrd1-Nab3-Sen1.

Nous avons ici montré un rôle global de l'interférence transcriptionnelle liée à la transcription pervasive antisens pour co-réguler un ensemble de gènes entre la phase exponentielle et la quiescence. Nous avons de plus mis en évidence l'existence d'une régulation mutuelle de la transcription antisens et de la transcription génique. Cette régulation seuil-dépendante permet un contrôle fin du niveau d'expression du gène selon les conditions extérieures. Elle coopère très probablement avec la régulation des facteurs de transcription, pour assurer une réponse transcriptionnelle rapide et efficace face aux variations environnementales.

De façon générale, la transcription antisens est donc souvent impliquée dans la répression des gènes non induits, afin d'éviter une fuite transcriptionnelle dans des conditions dans lesquelles ils sont normalement éteints.

Matériels et Méthodes

Souches et conditions de culture

Les souches utilisées proviennent de la collection de délétion Euroscarf, et dérivent des souches de référence BY4741 ou BY47412 de fond génétique MATa, ura3Δ0, his3Δ1, leu2Δ0, met15Δ0 ou MATα, ura3Δ0, his3Δ1, leu2Δ0, lys2Δ0 (Baker Brachmann *et al*, 1998). L'ensemble de ces souches est listé en Annexe 1.

Les cultures en phase exponentielles sont réalisées à 30°C, sous agitation, dans un milieu riche YPD. A une absorbance de 0,6-0,8, les cellules sont rincées à l'eau puis culottées par centrifugation et stockées à -80°C.

Les cellules quiescentes ont été purifiées par gradient de Percoll, selon un protocole publié en 2006 (Allen *et al*, 2006) : après l'ensemencement de milieu riche YPD, le milieu est naturellement appauvri en nutriments pendant 10 jours de culture à 30°C. Les cellules en phase stationnaire sont ensuite reprises dans du Tris-HCl 50 mM pH 7,5 et déposées sur un gradient de Percoll® (SIGMA). Ce gradient est obtenu par centrifugation d'une solution de NaCl 1,5M (1 volume) et Percoll® (9 volumes) à 19 250 G pendant 15 minutes à température ambiante (rotor à angle fixe). Les cellules migrent ensuite selon un gradient de densité dans le tube de gradient pendant une centrifugation à 400 G pendant 30 minutes à température ambiante (rotor à angle pivotant). Les cellules quiescentes sont notamment caractérisées par un épaississement de la membrane plasmique. Leur densité importante permet donc leur migration au fond du gradient. Ces cellules sont purifiées, reprises dans le tampon de Tris-HCl, puis culottées et stockées à -80°C.

Construction des souches pour les systèmes rapporteurs :

Les systèmes rapporteurs ont été construits sur deux vecteurs d'expression chez la levure (pRS315 et pRS313), par clonages successifs de produits de PCR ou de digestions plasmidiques (Amplification à l'ADN polymérase Phusion High-Fidelity de Thermo Fisher avec le protocole fourni, transformation de bactéries compétentes par électroporation). La présence de sites de restriction uniques à proximité des promoteurs géniques/antisens permet l'association du gène rapporteur avec des combinaisons de promoteurs différents. Pour cela, le vecteur a été ouvert au niveau du promoteur à remplacer, et co-transformé dans la levure avec une séquence contenant le nouveau promoteur et des régions d'homologie (par analogie

avec la technique du « Gap Repair », Lundblad & Struhl, 2008). L'ensemble des oligonucléotides utilisés sont listés en Annexe 2.

Le système rapporteur CUP a été construit dans le plasmide pRS315 et peut être intégré dans le génome de *S. cerevisiae* au niveau du locus CUP1 après son excision du plasmide, par digestion par EagI et SalI (Figure 27).

Le système rapporteur URA3-GFP a été construit dans le plasmide pRS313 (Figure 27).

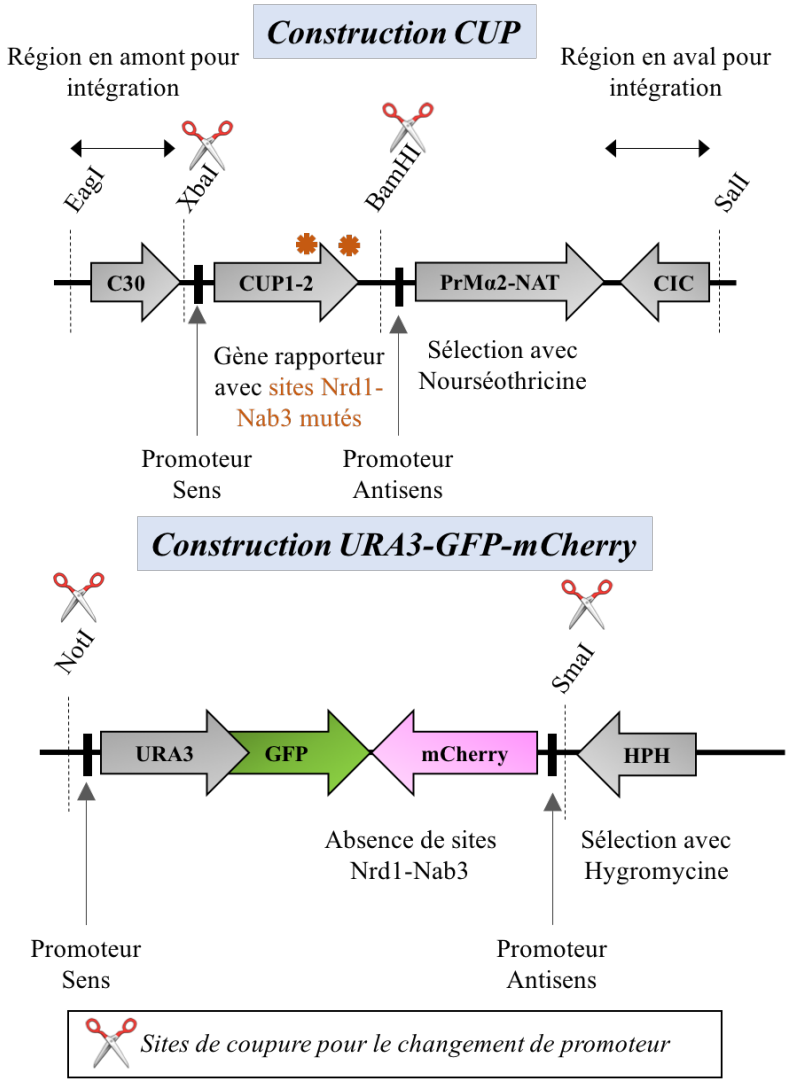


Figure 27 : Schéma des deux systèmes rapporteurs développés

Construction des souches pour les mutants NNS

La séquence de 37 pb correspondant au terminateur NNS a été insérée dans l'orientation sens ou antisens pour chaque gène, grâce à une méthode décrite précédemment (Erdeniz *et al*, 1997).

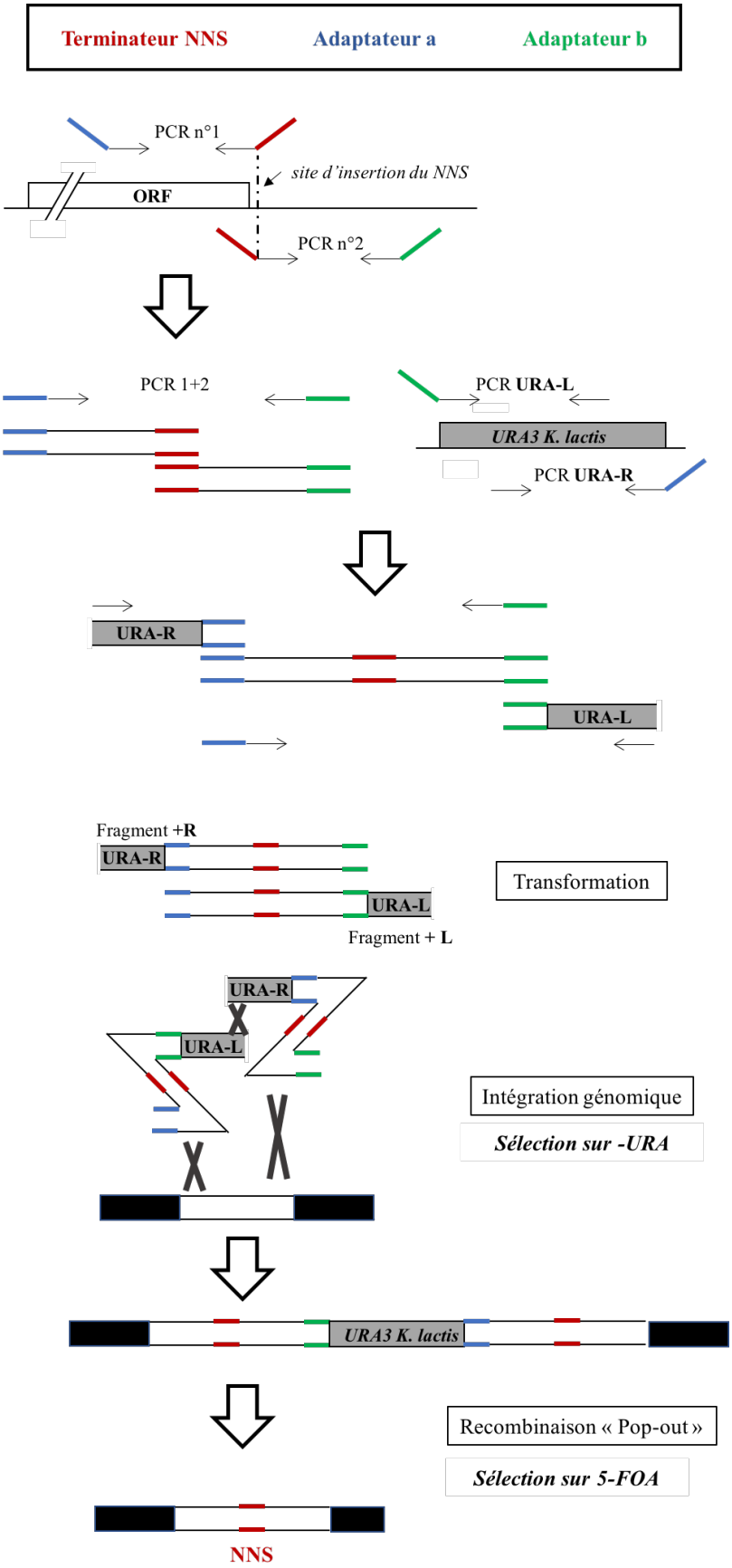


Figure 28 : Schéma de la méthode de mutagénèse utilisée pour l'insertion du NNS

Une reconstruction de PCR a permis de synthétiser le fragment correspondant à la séquence du gène avec l'insertion du NNS (PCR 1+2, Figure 28). Puis un fragment L ou R a été amplifié à partir du gène URA3 de la levure *Kluyveromyces lactis* et ajouté en amont ou en aval de la séquence modifiée (Fragments +L ou +R, Figure 28). Les fragments obtenus ont été co-transformés selon un protocole couramment utilisé (Gietz *et al*, 1992).

Le produit de transformation est étalé sur un milieu dépourvu d'uracile pour sélectionner l'intégration d'une séquence recombinée, contenant deux copies de la séquence modifiée et le gène URA3 reformé à partir des fragments +L et +R. Les clones sélectionnés sont ensuite criblés par PCR et les clones positifs sont repiqués sur milieu riche YPD pour une sélection ultérieure de la recombinaison « pop-out » sur 5-FOA (Figure 28). Le « pop-out » permet l'excision du gène URA3, et d'une copie de la séquence mutée, c'est à dire qu'il ne reste que l'insertion du NNS sans aucune autre modification de séquence. Les clones sélectionnés sont ensuite criblés par PCR pour l'insertion de 37 nucléotides, grâce à une courte amplification autour du site d'insertion et une migration sur un gel à 3% d'agarose (p/v). Enfin, les mutants NNS sont stockés, et utilisés pour la délétion du gène *UPF1* par amplification et transformation d'une cassette de délétion d'un mutant Euroscarf.

Extraction des ARN

Les ARN totaux ont été extraits en présence d'une solution dénaturante contenant notamment du Thiocyanate de Guanidium et du ß-mercaptoéthanol, selon le protocole décrit (Chomczynski & Sacchi, 2006). Leur concentration est ensuite mesurée par lecture d'absorbance à 260nm et les ARN sont dilués à une concentration d'environ 1µg/µl. 4µg de chaque échantillon sont ensuite déposés sur gel d'agarose (1%, p/v) pour vérifier l'absence de dégradation.

Hybridations par Northern-Blot

Les ARN (4µg) sont transférés sur membrane de nylon (Hybond N+, GE healthcare) selon le protocole recommandé par le fournisseur pour réaliser des hybridations avec des sondes radioactives. Les sondes ARN ont été synthétisées par transcription *in vitro* et marquées par incorporation de phosphate radioactif en utilisant le kit MAXIscript (ambion®) selon le protocole proposé. Les oligonucléotides utilisés comme sondes ADN ont été marqués par addition en 5' de phosphate radioactif catalysée par la T4 polynucleotide kinase (50 pmol

d'oligonucléotide, tampon de réaction A 1X, 5µl d'ATP γP32, 1µl de T4 polynucléotide kinase Thermo Scientific : incubation 30 min à 37°C et inactivation 5 min à 85°C).

PCR-quantitatives en temps réel

Afin d'éliminer d'éventuelles contaminations des préparations d'ARN totaux par de l'ADN, un traitement avec de la DNase I a été réalisé avec la TURBO™ DNase (ambion®) en suivant les recommandations du fournisseur. Après incubation pendant 30 minutes à 37°C, la DNase I est éliminée grâce à l'« Inactivation Reagent » fourni dans le kit. Le surnageant récupéré après 1 à 2 minutes de centrifugation est ensuite dosé par mesure de l'absorbance UV à 260 nm. 4µg d'ARN traités sont ensuite utilisés pour une transcription inverse (RT) en ADNc (2µg pour la RT-sens et 2µg pour la RT-antisens) en utilisant des oligonucléotides spécifiques à chaque gène testé et pour chaque orientation. La réaction de RT utilise la SuperScript® III (ThermoFisher) et est préparée selon le protocole mis à disposition par le fournisseur, en ajoutant de l'actinomycine D à 0,5mg/ml pour s'assurer de la directionnalité de la transcription réverse. Les produit de RT sont ensuite dilués au 1/10, et quatre dilutions en chaînes sont quantifiées par PCR quantitative grâce au qPCR Mix 2x Lo-Rox (Eurobiogreen).

Séquençage des ARN (RNA-seq)

Préparation des librairies de séquençage

Pour les librairies de séquençage, un « spike-in » de cellules en phase exponentielle de *S. pombe* a été ajouté aux culots cellulaires en quiescence et en phase exponentielle avant extraction d'ARN, pour permettre une normalisation future.

15µg d'ARN ont été traités à la TURBO™ DNase (ambion®) qui est ensuite éliminée grâce à l' « Inactivation Reagent » fourni dans le kit. Le surnageant récupéré après 1 à 2 minutes de centrifugation est ensuite dosé par mesure de l'absorbance UV à 260 nm. 5µg d'ARN sont déplétés en ARN ribosomiques, en utilisant le kit Ribo-Zéro Gold (Illumina®) selon le protocole fourni. Les librairies de RNA-seq ont été construites en utilisant le Kit Illumina Truseq Stranded mRNA et le protocole fourni. Puis des séquençages en "Single read 50" (SR50) ont été réalisés sur un Illumina Hiseq 2500 (Institut Pasteur, Plate-forme Transcriptomique PF2).

Traitement et analyse des données de séquençage

Les données de séquençage ont d'abord été traitées en supprimant les reads dupliqués, grâce à l'outil fqduplicate (ftp://ftp.pasteur.fr/pub/gensoft/projects/, version 1.1), puis les erreurs de séquence ont été corrigées en utilisant l'algorithme Musket (Liu et al, 2013; version 1.1). Les reads de mauvaise qualité ont été retirés grâce à fastq_qual_trimmer (https://github.com/ivars-

silamikelis/fastq_qual_trimmer, version 1.0), en choisissant un seuil de 20. Enfin, les adaptateurs Illumina ont été retirés avec Flexbar (Dodt et al, 2012; version 2.7). Les reads ont ensuite été alignés sur le génome en utilisant Bowtie (Langmead & Salzberg, 2012; version 2.2.3 avec les paramètres suivants : –N 1 –p 1 —no-unal –D 15 –R 2 –L 22 –I S,1,1.15) et une compilation des génomes de *S. cerevisiae* (séquence de référence S288C, version 64 obtenue à partir de la base de données « Saccharomyces Genome Database » (SGD) [http://www.yeastgenome.org/]) et de *S. pombe* (séquence de référence ASM294, version v2.19 obtenue à partir de la base de données PomBase [http://www.pombase.org/]) comme génomes de référence.

Après comptage des reads en utilisant HTSeq-count, les calculs d'expression différentielle des transcrits ont été faits par DESeq2 compris dans le pipeline SARTools (Varet et al, 2016; version 1.4.1). Les échantillons correspondant aux cellules exponentielles ou aux cellules quiescentes ont d'abord été normalisés dans des groupes séparés, en utilisant l'outil SARTools. Les facteurs de normalisation ont été extraits et utilisés pour la production de fichiers wig normalisés. Les échantillons quiescents ont été normalisés une seconde fois par rapport aux échantillons en phase exponentielle, grâce au spike-in de cellules de *S. pombe*. La médiane des reads comptés a été déterminée après la première étape de de normalisation, pour chaque échantillon. Puis une moyenne globale a été calculée pour tous les échantillons en phase exponentielle et tous les échantillons en quiescence. Un ratio phase exponentielle/quiescence a été calculé, et utilisé pour re-normaliser les fichiers wig des échantillons quiescents.

Le nombre de reads pour les transcrits antisens et les ARNm a été compté et visualisé dans une fenêtre de -50 à +200 nucléotides autour de chaque TSS, en utilisant le package « Counter RNAseq window » (CRAW) (voir http://bneron.pages.pasteur.fr/craw/).

Les données traitées ont été analysées en utilisant le logiciel R, puis les données décrites ont été déposées sur NCBI GEO avec le numéro d'accès GSE101368.

Observation des cellules par microscopie à fluorescence

Les expériences de microscopie à fluorescence ont été effectuées sur des cellules cultivées sur boîtes avec du milieu puis resuspendues dans de l'eau stérile et observées en lame et lamelles sous un microscope à épifluorescence Leica DMRXA. Les signaux fluorescents ont été collectés avec des filtres passe-bande pour l'excitation de la GFP pendant 2000 millisecondes et pour l'excitation de la mCherry pendant 300 millisecondes. L'acquisition des images est permise par une caméra CCD ORCAII-ER.

BIBLIOGRAPHIE

Alcid EA & Tsukiyama T (2016) Expansion of antisense lncRNA transcriptomes in budding yeast species since the loss of RNAi. *Nat. Struct. Mol. Biol.* **23:** 450–455

Allen C, Büttner S, Aragon AD, Thomas JA, Meirelles O, Jaetao JE, Benn D, Ruby SW, Veenhuis M, Madeo F & Werner-Washburne M (2006) Isolation of quiescent and nonquiescent cells from yeast stationary-phase cultures. *J. Cell Biol.* **174:** 89–100

Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S & Jacobson A (2004) A faux 3′-UTR promotes aberrant termination and triggers nonsense- mediated mRNA decay. *Nature* **432:** 112–118

Arigo JT, Carroll KL, Ames JM & Corden JL (2006) Regulation of Yeast NRD1 Expression by Premature Transcription Termination. *Mol. Cell* **21:** 641–651

Arribere JA & Gilbert WV (2013) Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Res.* **23:** 977–987

Baker Brachmann C, Davies A, Cost GJ, Caputo E, Li J, Hieter P & Boeke JD (1998) Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14:** 115–132

Battaglia S, Lidschreiber M, Baejen C, Torkler P, M Vos S & Cramer P (2017) RNA-dependent chromatin association of transcription elongation factors and Pol II CTD kinases. *eLife*

Berretta J, Pinskaya M & Morillon A (2008) A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in S. cerevisiae. *Genes Dev.* **22:** 615–626

Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH & Long RM (1998) Localization of ASH1 mRNA Particles in Living Yeast. *Mol. Cell* **2:** 437–445

Bird AJ, Gordon M, Eide DJ & Winge DR (2006) Repression of ADH1 and ADH3 during zinc deficiency by Zap1-induced intergenic RNA transcripts. *EMBO J.* **25:** 5726–5734

Brogna S & Wen J (2009) Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* **16:** 107

Bumgarner SL, Dowell RD, Grisafi P, Gifford DK & Fink GR (2009) Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc. Natl. Acad. Sci.* **106:** 18321–18326

Bumgarner SL, Neuert G, Voight BF, Symbor-Nagrabska A, Grisafi P, van Oudenaarden A & Fink GR (2012) Single-Cell Analysis Reveals that Noncoding RNAs Contribute to Clonal Heterogeneity by Modulating Transcription Factor Recruitment. *Mol. Cell* **45:** 470–482

Cairns BR (2009) The logic of chromatin architecture and remodelling at promoters. *Nature* **461:** 193–198

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

Camblong J, Iglesias N, Fickentscher C, Dieppois G & Stutz F (2007) Antisense RNA Stabilization Induces Transcriptional Gene Silencing via Histone Deacetylation in S. cerevisiae. *Cell* **131:** 706–717

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

Castelnuovo M & Stutz F (2015) Role of chromatin, environmental changes and single cell heterogeneity in non-coding transcription and gene regulation. *Curr. Opin. Cell Biol.* **34:** 16–22

Castelnuovo M, Zaugg JB, Guffanti E, Maffioletti A, Camblong J, Xu Z, Clauder-Munster S, Steinmetz LM, Luscombe NM & 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

Chomczynski P & Sacchi N (2006) The single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction: twenty-something years on. *Nat. Protoc.* **1:** 581

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

Colin J, Libri D & Porrua O (2011) Cryptic Transcription and Early Termination in the Control of Gene Expression. *Genet. Res. Int.* **2011:** e653494

Decourty L, Doyen A, Malabat C, Frachon E, Rispal D, Séraphin B, Feuerbach F, Jacquier A & Saveanu C (2014) Long Open Reading Frame Transcripts Escape Nonsense-Mediated mRNA Decay in Yeast. *Cell Rep.* **6:** 593–598

Dever TE, Kinzy TG & Pavitt GD (2016) Mechanism and Regulation of Protein Synthesis in Saccharomyces cerevisiae. *Genetics* **203:** 65–107

van Dijk EL, Chen CL, d'Aubenton-Carafa Y, Gourvennec 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

Egloff S & Murphy S (2008) Cracking the RNA polymerase II CTD code. *Trends Genet.* **24:** 280–288

Erdeniz N, Mortensen UH & Rothstein R (1997) Cloning-Free PCR-Based Allele Replacement Methods. *Genome Res.* **7:** 1174–1183

Feng J, Gan H, Eaton ML, Zhou H, Li S, Belsky JA, MacAlpine DM, Zhang Z & Li Q (2016) Noncoding Transcription Is a Driving Force for Nucleosome Instability in spt16 Mutant Cells. *Mol. Cell. Biol.* **36:** 1856–1867

Fox MJ, Gao H, Smith-Kinnaman WR, Liu Y & Mosley AL (2015) The Exosome Component Rrp6 Is Required for RNA Polymerase II Termination at Specific Targets of the Nrd1-Nab3 Pathway. *PLOS Genet.* **11:** e1004999

Gietz D, St Jean A, Woods RA & Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.* **20:** 1425

Gu M, Naiyachit Y, Wood TJ & Millar CB (2015) H2A.Z marks antisense promoters and has positive effects on antisense transcript levels in budding yeast. *BMC Genomics* **16:** 99

Gudipati RK, Neil H, Feuerbach F, Malabat C & Jacquier A (2012) The yeast RPL9B gene is regulated by modulation between two modes of transcription termination. *EMBO J.* **31:** 2427–2437

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

Hahn S (2004) Structure and mechanism of the RNA Polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* **11:** 394–403

Hahn S & Young ET (2011) Transcriptional Regulation in Saccharomyces cerevisiae: Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators. *Genetics* **189:** 705–736

Hainer SJ, Charsar BA, Cohen SB & Martens JA (2012) Identification of Mutant Versions of the Spt16 Histone Chaperone That Are Defective for Transcription-Coupled Nucleosome Occupancy in Saccharomyces cerevisiae. *G3 GenesGenomesGenetics* **2:** 555

Hainer SJ, Pruneski JA, Mitchell RD, Monteverde RM & Martens JA (2011) Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev.* **25:** 29–40

Han Y, Yan C, Nguyen THD, Jackobel AJ, Ivanov I, Knutson BA & He Y (2017) Structural mechanism of ATP-independent transcription initiation by RNA polymerase I. *eLife* **6:**

Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5489313/ [Accessed August 2, 2017]

He F & Jacobson A (2015) Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. *Annu. Rev. Genet.* **49:** 339–366

Heidemann M, Hintermair C, Voß K & Eick D (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1829:** 55–62

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. *Mol. Cell* **32:** 685–695

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

Jacquier A (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat. Rev. Genet.* **10:** 833–844

Jensen TH, Jacquier A & Libri D (2013) Dealing with Pervasive Transcription. *Mol. Cell* **52:** 473–484

Jonkers I & Lis JT (2015) Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16:** 167–177

Kaplan CD, Laprade L & Winston F (2003) Transcription Elongation Factors Repress Transcription Initiation from Cryptic Sites. *Science* **301:** 1096–1099

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

Kim K-Y & Levin DE (2011) Mpk1 MAPK Association with the Paf1 Complex Blocks Sen1-Mediated Premature Transcription Termination. *Cell* **144:** 745–756

Kim T, Xu Z, Clauder-Münster S, Steinmetz LM & Buratowski S (2012) Set3 HDAC Mediates Effects of Overlapping Noncoding Transcription on Gene Induction Kinetics. *Cell* **150:** 1158–1169

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

Kuroha K, Tatematsu T & Inada T (2009) Upf1 stimulates degradation of the product derived from aberrant messenger RNA containing a specific nonsense mutation by the proteasome. *EMBO Rep.* **10:** 1265–1271

Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, Chu A, Davis RW,

Steinmetz LM & Primig M (2011) Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc. Natl. Acad. Sci.* **108:** 1058–1063

Lee JT (2010) The X as Model for RNA's Niche in Epigenomic Regulation. *Cold Spring Harb. Perspect. Biol.* **2:** a003749

Lenstra TL, Coulon A, Chow CC & Larson DR (2015) Single-Molecule Imaging Reveals a Switch between Spurious and Functional ncRNA Transcription. *Mol. Cell* **60:** 597–610

Libri D (2015) Endless Quarrels at the End of Genes. *Mol. Cell* **60:** 192–194

Lin C-H, Tsai ZT-Y & Wang D (2013) Role of antisense RNAs in evolution of yeast regulatory complexity. *Genomics* **102:** 484–490

Lundblad V & Struhl K (2008) Yeast. In *Current Protocols in Molecular Biology*, Ausubel FM Brent R Kingston RE Moore DD Seidman JG Smith JA & Struhl K (eds) Hoboken, NJ, USA: John Wiley & Sons, Inc. Available at: http://doi.wiley.com/10.1002/0471142727.mb1300s82 [Accessed March 14, 2017]

Luse DS (2013) Promoter clearance by RNA polymerase II. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **1829:** 63–68

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

Marquardt S, Escalante-Chong R, Pho N, Wang J, Churchman LS, Springer M & Buratowski S (2014) A Chromatin-Based Mechanism for Limiting Divergent Noncoding Transcription. *Cell* **157:** 1712–1723

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

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

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

Matera AG, Terns RM & Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.* **8:** 209

Meinhart A, Kamenski T, Hoeppner S, Baumli S & Cramer P (2005) A structural perspective of CTD function. *Genes Dev.* **19:** 1401–1415

Millar CB & Grunstein M (2006) Genome-wide patterns of histone modifications in yeast. *Nat. Rev. Mol. Cell Biol.* **7:** 657–666

Mitchell P & Tollervey D (2003) An NMD Pathway in Yeast Involving Accelerated

Deadenylation and Exosome-Mediated 3′→5′ Degradation. *Mol. Cell* **11:** 1405–1413

Murray SC, Haenni S, Howe FS, Fischl H, Chocian K, Nair A & Mellor J (2015) Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic Acids Res.* **43:** 7823–7837

Murray SC, Serra Barros A, Brown DA, Dudek P, Ayling J & Mellor J (2012) A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic Acids Res.* **40:** 2432–2444

Nadal-Ribelles M, Solé C, Xu Z, Steinmetz LM, de Nadal E & 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 LM & Jacquier A (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457:** 1038–1042

Nguyen T, Fischl H, Howe FS, Woloszczuk R, Barros AS, Xu Z, Brown D, Murray SC, Haenni S, Halstead JM & others (2014) Transcription mediated insulation and interference direct gene cluster expression switches. *Elife* **3:** e03635

Pelechano V & Steinmetz LM (2013) Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14:** 880–893

Pinskaya M, Gourvennec S & Morillon A (2009) H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J.* **28:** 1697–1707

Porrua O, Boudvillain M & Libri D (2016) Transcription Termination: Variations on Common Themes. *Trends Genet.* **32:** 508–522

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

Porrua O & Libri D (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16:** 190–202

Radonjic M, Andrau J-C, Lijnzaad P, Kemmeren P, Kockelkorn TTJP, van Leenen D, van Berkum NL & Holstege FCP (2005) Genome-Wide Analyses Reveal RNA Polymerase II Located Upstream of Genes Poised for Rapid Response upon S. cerevisiae Stationary Phase Exit. *Mol. Cell* **18:** 171–183

Rando OJ & Winston F (2012) Chromatin and Transcription in Yeast. *Genetics* **190:** 351–387

Rhee HS & Pugh BF (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483:** 295–301

Sainsbury S, Bernecky C & Cramer P (2015) Structural basis of transcription initiation by

RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **16:** 129–143

Schlissel MS (2003) Regulating antigen-receptor gene assembly. *Nat. Rev. Immunol.* **3:** 890–899

Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J & Cramer P (2013) Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell* **155:** 1075–1087

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

Smolle M, Venkatesh S, Gogol MM, Li H, Zhang Y, Florens L, Washburn MP & Workman JL (2012) Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat. Struct. Mol. Biol.* **19:** 884–892

Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM & Proudfoot NJ (2012) Gene Loops Enhance Transcriptional Directionality. *Science* **338:** 671–675

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

Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J & Libri D (2006) Transcription Termination and Nuclear Degradation of Cryptic Unstable Transcripts: A Role for the Nrd1-Nab3 Pathway in Genome Surveillance. *Mol. Cell* **23:** 853–864

Tisseur M, Kwapisz M & Morillon A (2011) Pervasive transcription – Lessons from yeast. *Biochimie* **93:** 1889–1896

Tudek A, Porrua O, Kabzinski T, Lidschreiber M, Kubicek K, Fortova A, Lacroute F, Vanacova S, Cramer P, Stefl R & Libri D (2014) Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Mol. Cell* **55:** 467–481

Turowski TW, Leśniewska E, Delan-Forino C, Sayou C, Boguta M & Tollervey D (2016) Global analysis of transcriptionally engaged yeast RNA polymerase III reveals extended tRNA transcripts. *Genome Res.* **26:** 933–944

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

Uwimana N, Collin P, Jeronimo C, Haibe-Kains B & Robert F (2017) Bidirectional terminators in Saccharomyces cerevisiae prevent cryptic transcription from invading neighboring genes. *Nucleic Acids Res.* **45:** 6417–6426

van Werven FJ, 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

Venkatesh S, Li H, Gogol MM & Workman JL (2016) Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nat. Commun.* **7:** 13610

Venkatesh S & Workman JL (2013) Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development: Regulation of transcription elongation and implications. *Wiley Interdiscip. Rev. Dev. Biol.* **2:** 685–700

Venters BJ & Pugh BF (2008) A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome Res.* **19:** 360–371

Vera JM & Dowell RD (2016) Survey of cryptic unstable transcripts in yeast. *BMC Genomics* **17:** 305

Wery M, Descrimes M, Vogt N, Dallongeville A-S, Gautheret D & Morillon A (2016) Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Mol. Cell* **61:** 379–392

Whitehouse I, Rando OJ, Delrow J & Tsukiyama T (2007) Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450:** 1031–1035

Woo H, Dam Ha S, Lee SB, Buratowski S & Kim T (2017) Modulation of gene expression dynamics by co-transcriptional histone methylations. *Exp. Mol. Med.* **49:** e326

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

Xu Z, Wei W, Gagneur J, Clauder-Münster S, Smolik M, Huber W & Steinmetz LM (2011) Antisense expression increases gene expression variability and locus interdependency. *Mol. Syst. Biol.* **7:** 468

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W & Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457:** 1033–1037

Yassour M, Pfiffner J, Levin JZ, Adiconis X, Gnrike A, Nusbaum C, Thompson D-A, Friedman N & Aviv R (2010) Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol.*

Yu Y, Yarrington RM, Chuong EB, Elde NC & Stillman DJ (2016) Disruption of promoter memory by synthesis of a long noncoding RNA. *Proc. Natl. Acad. Sci.* **113:** 9575–9580

Annexe 1 : Liste des souches utilisées

LMA ID	ORF	Type	Ploidie	Marker	Original Strain
LMA1057	YEL063C	ΔYEL063C	Mat a	CAN	BY4741
LMA1790-A	YMR080C /YEL063C	Δnam7 : KAN / Δcan1	Mat a	KAN, CAN	LMA1676/LMA1668
LMA1676	YOR001W/YEL063C	Δrrp6 : hph / Δcan1	Mat a	HPH , CAN	LMA1617
LMA1772-A	YMR080C/YEL063C/YOR001W	Δnam7 : KAN / Δrrp6 : hph / Δcan1	Mat a	KAN, HPH, CAN	LMA1676/LMA1668
LMA3198	YKR046C	PET10:: Terminateur NNS Antisens	Mat a	None	BY4741
LMA4686	YKR046C	Δpet10 : hph	2n	HYGRO	BY4741/LMA4142
LMA4240	YDR380W	ARO10:: Terminateur NNS Antisens	Mat alpha	None	BY4742
LMA4853	YMR118C	SHH3:: Terminateur NNS Antisens	Mat a	None	BY4742
LMA4854	YBL049W	MOH1:: Terminateur NNS Antisens	Mat alpha	None	BY4742
LMA4856	YGR110W	CLD1:: Terminateur NNS Antisens	Mat alpha	None	BY4742
LMA4054	YER055C	HIS1:: Terminateur NNS Sens	Mat alpha	None	BY4742
LMA4106	YMR118C	SHH3:: Terminateur NNS Sens	Mat alpha	None	BY4742
LMA4109	YKR046C	PET10:: Terminateur NNS Sens	Mat alpha	None	BY4742

LMA4243	YDR380W	ARO10:: Terminateur NNS Sens	Mat alpha	None	BY4742
LMA4855	YBL049W	MOH1:: Terminateur NNS Sens	Mat alpha	None	BY4742
LMA4857	YGR110W	CLD1:: Terminateur NNS Sens	Mat alpha	None	BY4742
LMA4142	YKR046C	Δpet10 : hph	Mat alpha	HYGRO	BY4742
LMA4215	YKR046C/YMR080C	Δnam7 : KAN / PET10:: Terminateur NNS Antisens	Mat a	KAN	LMA3198
LMA4687	YKR046C	PET10 NNS antisense, PET10 NNS sense	2n	None	LMA3198/LMA4109
LMA4684	YKR046C	PET10:: Terminateur NNS Antisens, PET10::KAN	2n	KAN	LMA3198/LMA4142
LMA4152	YER055C/YMR080C	Δnam7 : KAN / HIS1:: Terminateur NNS Sens	Mat alpha	KAN	LMA4054
LMA4225	YMR118C/YMR080C	Δnam7 : KAN / SHH3_Sense NNS terminator	Mat alpha	KAN	LMA4106
LMA4228	YKR046C/YMR080C	Δnam7 : KAN / PET10_Sense NNS terminator	Mat alpha	KAN	LMA4109
LMA4273	YDR380W /YMR080C	Δnam7 : KAN/ ARO10:: Terminateur NNS Antisens	Mat alpha	KAN	LMA4240
LMA4276	YDR380W /YMR080C	Δnam7 : KAN/ ARO10::	Mat alpha	KAN	LMA4243

		Terminateur NNS Sens			
LMA4858	YMR118C/YMR080C	Δnam7 : KAN /SHH3:: Terminateur NNS Antisens	Mat alpha	KAN	LMA4853
LMA4859	YBL049W/YMR080C	Δnam7 : KAN /MOH1:: Terminateur NNS Antisens	Mat alpha	KAN	LMA4854
LMA4860	YBL049W/YMR080C	Δnam7 : KAN /MOH1:: Terminateur NNS Sens	Mat alpha	KAN	LMA4855
LMA4861	YGR110W/YMR080C	Δnam7 : KAN /CLD1:: Terminateur NNS Antisens	Mat a	KAN	LMA4856
LMA4862	YGR110W/YMR080C	Δnam7 : KAN /CLD1:: Terminateur NNS Sens	Mat a	KAN	LMA4857

Annexe 2 : liste des oligonucléotides utilisés

Nom	Description	Séquence
GB_909	PET10_fwd	CAACGCTTCACCACTGTC
GB_988	PET10_URA3_fwd	CAAGAGAAGACCAAACCAATTCTAAGCCCGCGGCTGTGTCGAC CAATTAGTTTGTGAGTTTAGTATACATGC
GB_989	PET10_URA3_rev	AGAGTGGCAGGAAAAAAAATCAGGCGCACGATTAGCGCAAAA CCAAATTACAGATCTGGCTTTTCAATTC
GB_990	PET10_NNS_AS_rev	TAAGTCTTGATATATAACAATTAGCTTGCTAATTGGTCGACACA GCC
GB_991	NNS_AS_PET10_fwd	GTTATATATCAAGACTTAATTCATTACTAATTTGGTTTTGCGCT AATCGTGC
GB_992	PET10 _rev	CTTTGGTATTCAATTGTATATAGTGC
AC-466	A_SHH3__fwd	aattccagctgaccaccatgCAAAGAATTCCACAATGGCC
AC-467	SHH3_NNS-AS-rev	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGCCAGATG AGATGCCTTATACC
AC-468	NNS-AS_SHH3_fwd	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACGATATGG GCTACGAATTGAC
AC-469	B_SHH3__rev	gatccccgggaattgccatgGCTCTTGCGTTCCTTCC
AC-470	SHH3_vérif	GCCGGCACACTCGTCT
AC-471	A_MOH1__fwd	aattccagctgaccaccatgGTAACACTTAAGTGGAGAGCAGAG
AC-472	MOH1_NNS-AS-rev	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGCATCATT GCTGCTCTGCAAG
AC-473	NNS-AS_MOH1_fwd	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACATCAGCA GTATAAGGAAGGAAAGTTTATC
AC-474	B_MOH1__rev	gatccccgggaattgccatgCGTATTACAACATGGGAAACACGAC
AC-475	MOH1_vérif	TGCAGAGCAGCAATGATGATC
AC-476	MOH1_vérif	CGAGAGGTGCTAGATAGCAGAC
AC-508	A_CLD1_AS__fwd	aattccagctgaccaccatgCAACAATACTGCTTCAGAGGTTG
AC-509	CLD1_NNS-AS-rev	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGCCACCAT TCACATCCATCCAATC
AC-510	NNS-AS_CLD1_fwd	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACACTTAGA GTATCAAGATTCTTGAAAGAGAAG
AC-511	B_CLD1_AS__rev	gatccccgggaattgccatgCTCCATAAAGACAGTGGTTCTATTGG
AC-512	CLD1_AS_vérif	AGTAATTGTGACTGGCTTTGGC
AC-513	CLD1_AS_vérif	CCTGAATGAGGAACAATGATGACG
AC_393	A_ARO10	aattccagctgaccaccatgCAATGCGGTTAGAAGATCCTAC
AC_394	ARO10_B	gatccccgggaattgccatgCGAAGTCTCCGAATGCTTC

AC_395	ARO10_rev_NNS antisens	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGCATGGCAATTCCAACACCTAG
AC_396	NNS antisens_ARO10_fwd	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACCAAGACCACTCAAACGCTC
AC_397	vérif ARO10	AACCCTGGTGATGTTGTCG
AC_398	vérif ARO10	GCTGCACCGTCACCTTC
AC_216	A_PET10	aattccagctgaccaccatgCTGCCGAAGAGCAAGACG
AC_217	PET10_B	gatccccgggaattgccatgCTGGTCGTTGAAGACAACTCTG
AC_218	PET10_NNS sens_rev	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACCGACCAGGTTGCTTGTGG
AC_219	NNS sens_PET10_fwd	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGCATCTGCAAAGATACCCAGC
AC_222	PET10_verif	CCATCTGTGGAATTGCC
AC_223	PET10_verif	GTGGCAGAGATTCCGC
AC_224	A_SHH3	aattccagctgaccaccatgCATGTTTGTACCAGAAGTACTTGC
AC_225	SHH3_B	gatccccgggaattgccatgCGTGATACCACTTGTTTAGATCC
AC_226	SHH3_NNS sens_rev	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACTATTCTTACGCTTCTGAATATTTCTGAG
AC_227	NNS sens_SHH3_fwd	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGATGAAAGCAACCATTCAAAGAGTAAC
AC_228	SHH3_verif	CGTTATCCTATCACATTATCAAATGAC
AC_229	SHH3_verif	GCCATTGTGGAATTCTTTG
AC-477	A_MOH1_Sense__fwd	aattccagctgaccaccatgGCAAGAGCGAAGTTGAAGCC
AC-478	MOH1_NNS-S-rev	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACAAGATAAGCTTCTCCCTTTTCCTTC
AC-479	NNS-S_MOH1_fwd	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGACTGTAGAAGAAAATGGGATTGCG
AC-480	B_MOH1_Sense__rev	gatccccgggaattgccatgCGAGAGGTGCTAGATAGCAGAC
AC-481	MOH1_Sense_vérif	GATGATGGGGAAGATAACGG
AC-482	MOH1_Sense_vérif	GAGAGCAGAGAGTAGCAGC
AC-514	A_CLD1_S__fwd	aattccagctgaccaccatgGCAGATTAGACTTGGGAGTTATCAC
AC-515	CLD1_NNS-S-rev	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACGTAACATCTCTTGCGGAGGCTC
AC-516	NNS-S_CLD1_fwd	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGCAGAGAAACAGATATTAAGAACTCCGTCAG
AC-517	B_CLD1_S__rev	gatccccgggaattgccatgGGCATGTATGCACCACTCGTTATC
AC-518	CLD1_S_vérif	GGAAACAAGCTTGGGACTCTATAG
AC_328	A_ARO10	aattccagctgaccaccatgCGATGCTTGTACACCTCATGTAG
AC_329	ARO10_B	gatccccgggaattgccatgATGTAGGTTCCGATCACTAAAGTTAC

AC_330	ARO10_rev_NNS sens	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACTTGAGTATCAAGTTGAGAGGGTTG
AC_331	NNS sens_ARO10_fwd	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGAACAAGTTGACGCGACTTCTG
AC_332	vérif ARO10	GCGTACAACAACGTCTTAGC
AC_333	vérif ARO10	CCGAGCAAGCGACTCTATC
AC_334	vérif ARO10	GTATGTAATAGGTTAGTGGCATC
AC_335	vérif ARO10	TTGTAACAGGTGCCATGC
AC_161	A_HIS1	AATTCCAGCTGACCACCATGGCTTCCAAGTGAGTCACC
AC_213	HIS1_B	GATCCCCGGGAATTGCCATGGCTTGTCTTCAGGTGCG
AC_193	HIS1_rev_NNS sens	CAAGCTAATTGTTATATATCAAGACTTAATTCATTACGGTCAACACCAGTTATACC
AC_194	NNS sens_HIS1_fwd	GTAATGAATTAAGTCTTGATATATAACAATTAGCTTGAAGTTCGTGAATCTAACGTC
CS_1199	B_KlURA3_fwd	CATGGCAATTCCCGGGGATCGTGATTCTGGGTAGAAGATCG
CS_1200	int3prime_KlURA3_rev	GAGCAATGAACCCAATAACGAAATC
CS_1201	int5prime_KlURA3_fwd	CTTGACGTTCGTTCGACTGATGAGC
CS_1202	A_KlURA3_rev	CATGGTGGTCAGCTGGAATTCGATGATGTAGTTTCTGGTT
GB_941	PET10 Sens Northern	CTGTGAACCAGTGGGTTTCC
GB_944	PET10 Sens Northern	TAATACGACTCACTATAGGGCTTGCAAGTGCAATGGCTTG
GB_942	PET10 Antisens Northern	TAATACGACTCACTATAGGGCTGTGAACCAGTGGGTTTCC
GB_943	PET10 Antisens Northern	CTTGCAAGTGCAATGGCTTG
GB_1027	SHH3 Sens Northern	GCAACCATTCAAAGAGTAAC
GB_1028	SHH3 Sens Northern	TAATACGACTCACTATAGGGCCCACGGTTGGTCAATTC
GB_1037	SHH3 Antisens Northern	TAATACGACTCACTATAGGGCAACCATTCAAAGAGTAAC
GB_1038	SHH3 Antisens Northern	CCCACGGTTGGTCAATTC
GB_1274	MOH1 Sens Northern	GAAAATGGGATTGCGTTACTCC
AC_495	MOH1 Sens Northern	TAATACGACTCACTATAGGGGCAGACAAAGGAAGACATCAAGTAC
AC_496	MOH1 Antisens Northern	TAATACGACTCACTATAGGGCGATGACTTAGAGAACAAAGCAG
AC_497	MOH1 Antisens Northern	CAGCTAGTTTGCAGTTCGTAC
AC_506	CLD1 Sens Northern	TAATACGACTCACTATAGGGGCATGTATGCACCACTCGTTATC
AC_507	CLD1 Sens Northern	GCTGACTCATCGAATACACGG

AC_504	CLD1 Antisens Northern	TAATACGACTCACTATAGGGCAACAATACTGCTTCAGAGGTTG
AC_505	CLD1 Antisens Northern	CCACCATTCACATCCATCCAATC
AC_374	ARO10 Sens Northern	ACAAGTTGACGCGACTTCTG
AC_375	ARO10 Sens Northern	TAATACGACTCACTATAGGGGCATAAGCGGCGTTCAGTTC
AC_372	ARO10 Antisens Northern	TAATACGACTCACTATAGGGGAACAATAACGGCTACACTATTG
AC_373	ARO10 Antisens Northern	CGAACAATTGGTAGCAGTG
AC_175	HIS1 Sens Northern	ATGGATTTGGTGAACCATC
AC_176	HIS1 Sens Northern	TAATACGACTCACTATAGGGGCCATTTACGGGGACTTG
AC_173	HIS1 Antisens Northern	TAATACGACTCACTATAGGGATGGATTTGGTGAACCATC
AC_174	HIS1 Antisens Northern	GCCATTTACGGGGACTTG
GB_987	SCR1 Northern	CAATTATCCGACTGATATGTGC
FF_3033	CPS1 RT Sens & Antisens	TGACCTGGGACATCCACG
AC_407	PET10 RT Sens	GCTTCACTATGGCGTCGTAC
AC_429	PET10 RT Sens	CAGACTGCAACTTGTCAGTTG
AC_500	SHH3 RT Sens	GCCGGCTAAAACGATTGATCCG
GB_1038	SHH3 RT Sens	CCCACGGTTGGTCAATTC
AC_430	PET10 RT Antisens	GGGCCTTCAAGTTGGACGAG
AC_431	PET10 RT Antisens	AGTTGTTGCTGCTGGGAAGG
AC_63	SHH3 RT Antisens	GCGACTGTGATAGCTG
AC_414	CPS1 qPCR	ACCAGTAGTCACATAGAACTCTTCG
AC_415	CPS1 qPCR	AGAACGGTGACGATTACATTATCC
AC_406	PET10 qPCR	ACGGCTTGCTGTACAATTGG
AC_428	PET10 qPCR	CGATCTCCTTCGATGCGGTAGAAC
AC_118	SHH3 qPCR	CTGGCTCTTGGATTCTATGCTTTCAC
AC_119	SHH3 qPCR	CAGATATGCTGCAGAGCC
AC_1	RSC30	GTAGTAATGTTCATGGAAACACTATTGTGAGAAGCGATAGTCC AGATGTG
AC_2	CUP1_ NNS*	GTATggatccTCATTTCCCAGAGCAGCATGACTTTTTGGTTTCCTCA GACTTGTTACC
AC_3	Terminateur His3	TGAggatccATACGATATATATACATGTGTATATATGTATACCT
AC_4	Terminateur His3_CIC1	CAGCAAATAGTTAGATGAATATATTAAAGACTATTCGTTcccggg AATGACACGTATAGAATGATGC
AC_5	Terminateur His3_CIC1	GGTAATGCATCATTCTATACGTGTCATTcccgggAACGAATAGTCT TTAATATATTCATCTAACTATTTGC

AC_7	CIC1	ATTcccgggAACGAATAGTCTTTAATATATTCATCTAAC
AC_8	CIC1	GAGgtcgacGGTTTACATGAATCAGTTGCC
AC_9	Prom ADE2 - *URA3*	GTGgcggccgcCCAcACGaGAAACCTGTCG
AC_17	GFP	GTTCTTCTCCTTTACTGttcatTCTAGA*TAGAAGACCCACTAT*GTTTTGCTGGCCGCATCTTCTC
AC_11	URA3-GFP	GAGAAGATGCGGCCAGCAAAACATAGTGGGTCTTCTAATGGTGAGCAAGGGCG
AC_12	GFP	TTAggatccCTTGTACAGCTCGTCCATGC
AC_13	mCherry	AAGggatccTAAGCCTAGGATTCCcttgtacagctcgtccatgccg
AC_14	mCherry	CTAATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTatggtgagcaagggcgagg
AC_15	Terminateur Ura3	cctcgcccttgctcaccatAAAACTGTATTATAAGTAAATGCATGTATACTAAACTCACAAATTAGAG
AC_16	Terminateur Ura3	atcgaattcCCCGGGTAATAACTGATATAATTAAATTGAAGCTCTAATTTGTGAGTTTAGTATAC
AC_18	URA3-GFP	GAGAAGATGCGGCCAGCAAAAC*ATAGTGGGTCTTCTA*TCTAGAatgaaCAGTAAAGGAGAAGAAC
AC_26	CIC1	GGGCAACTGATTCATGTAAACCgtcgacCTC
AC_27	Promoteur de TPK1_Construction 1	CAGGCTGATATCTTAGCCTTGTTACTAGTTAGAAAAAGACATTTTTGCTGTTACACCACATATTTTTAGTTACTTG
AC_28	Promoteur de TPK1_Construction 1	CATTTTATGTGATGATTGATTGATTGATTGTACAGTTTGTTTTTCTTAATATCTATTTTAGCTTTATCAAATTAAAACTCTTAGATGG
AC_29	Promoteur Antisens de TPK1_Construction 1	GCGGTAACAAGTCTGAGGAAACCAAAAAGTCATGCTGCTCTGGGAAATGACACTGTTCTTGTAGATTGCATAC
AC_30	Promoteur Antisens de TPK1_Construction 1	GATTCGATACTAACGCCGCCATCCAGTGTCGAAAACGAGCTCGAATTCATCGATGATCCAATTGCCCTTTATTTGG
AC_31	Promoteur Antisens de TPK1_Construction 2	catgaactccttgatgatggccatgttatcctcctcgcccttgctcaccatCACTGTTCTTGTAGATTGCATAC
AC_32	Promoteur de TPK1_Construction 2	gtaatacgactcactatagggcgaattggagctccaccgcggtggcggccgcTACACCACATATTTTTAGTTACTTG
AC_33	Promoteur de TPK1_Construction 2	CAACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACATTTAGCTTTATCAAATTAAAACTCTTAGATGG
AC_34	Promoteur de CPS1_Construction 1	CAGGCTGATATCTTAGCCTTGTTACTAGTTAGAAAAAGACATTTTTGCTGTACTGTTATCCTAAATGCTCAATG
AC_35	Promoteur de CPS1_Construction 1	CATTTTATGTGATGATTGATTGATTGATTGTACAGTTTGTTTTTCTTAATATCTATTTTAATGTGATGATGCAGGATTAG
AC_36	Promoteur Antisens de CPS1_Construction 1	GCGGTAACAAGTCTGAGGAAACCAAAAAGTCATGCTGCTCTGGGAAATGACTATGTGACTACTGGTTTATTCTC
AC_37	Promoteur Antisens de	GATTCGATACTAACGCCGCCATCCAGTGTCGAAAACGAGCTCG

	CPS1_Construction 1 et 2	AATTCATCGATGATTTTAAGTCATTGTATGGTCATATTATC
AC_38	Promoteur Antisens de CPS1_Construction 2	catgaactccttgatgatggccatgttatcctcctcgcccttgctcaccatCTATGTGACTACTGGTTTATTCTC
AC_39	Promoteur de CPS1_Construction 2	gtaatacgactcactatagggcgaattggagctccaccgcggtggcggccgcACTGTTATCCTAAATGCTCAATG
AC_40	Promoteur de CPS1_Construction 2	CAACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACATTTAATGTGATGATGCAGGATTAG
AC_41	Promoteur de PDR17_Construction 1	CAGGCTGATATCTTAGCCTTGTTACTAGTTAGAAAAAGACATTTTTGCTGTCATAAATTTTCTATTTACTTTTCTTATTTTTCC
AC_42	Promoteur de PDR17_Construction 1	CATTTTATGTGATGATTGATTGATTGATTGTACAGTTTGTTTTTCTTAATATCTATTTGTTCTTCTTCTTGTGCTTTC
AC_43	Promoteur Antisens de PDR17_Construction 1	GCGGTAACAAGTCTGAGGAAACCAAAAAGTCATGCTGCTCTGGGAAATGAAACGTTTCCCGCTACAC
AC_44	Promoteur Antisens de PDR17_Construction 1	GATTCGATACTAACGCCGCCATCCAGTGTCGAAAACGAGCTCGAATTCATCGATGATATTAGCTGGAACTAATAGTATATATAATTTTC
AC_45	Promoteur Antisens de PDR17_Construction 2	catgaactccttgatgatggccatgttatcctcctcgcccttgctcaccatAACGTTTCCCGCTACAC
AC_46	Promoteur de PDR17_Construction 2	gtaatacgactcactatagggcgaattggagctccaccgcggtggcggccgcCATAAATTTTCTATTTACTTTTCTTATTTTTCC
AC_47	Promoteur de PDR17_Construction 2	CAACAGGACTAGGATGAGTAGCAGCACGTTCCTTATATGTAGCTTTCGACATTGTTCTTCTTCTTGTGCTTTC