Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche
researchportal.unamur.be
University of Namur

THESIS / THÈSE

DOCTOR OF SCIENCES

New insights into the roles of the phosphorylation of serine 2 within the RNA polymerase II CTD

Yague-Sanz, Carlo

Award date:
2018

Awarding institution:
University of Namur

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 02. May. 2026

Carlo Yague-Sanz
OCTOBER 2018 | THE UNIVERSITY OF NAMUR

Doctoral thesis

NEW INSIGHTS INTO THE ROLES OF THE PHOSPHORYLATION OF SERINE 2 WITHIN THE RNA POLYMERASE II CTD.

THESIS DIRECTOR: Damien Hermand.

JURY: Marielle Boonen, Michel Werner, Dominique Helmlinger & Jean-Pierre Gillet

TABLE OF CONTENTS

Acknowledgments | Remerciements **7**

SHORT ENGLISH VERSION OF THE ACKNOLEDGMENTS *11*

How to read this thesis **12**

Abbreviations **13**

General Introduction **15**

1. Life & Transcription **15**

2. **The fission yeast *S. pombe*** **18**

3. The RNA Pol II C-terminal domain **21**

A. Complexity of the Pol II C-terminal domain (CTD) *21*

B. Pol II CTD coordinates mRNA maturation with transcription *23*

C. Pol II CTD coordinates chromatin remodeling with transcription *26*

D. Gene- or condition-specific roles for CTD modifications *28*

4. General objectives of this study **29**

Part 1: Phosphorylation of Ser2 on the RNA polymerase II CTD opposes tRNA expression. **31**

1. Introduction **31**

A. The Pol III transcriptional machinery *31*

B. Pol III regulation *33*

C. Maturation of pre-tRNAs. *34*

D. Interplay between Pol II and Pol III transcription *36*

E. Aim of the work *36*

2. Results & Discussion **37**

A. A global view on transcriptional changes upon Ser2-P loss. *37*

B. Genetic interactions connect Ser2-P of the RNA polymerase II with Pol III transcription *45*

C. Ser2-P opposes tRNA expression in a Maf1-independent manner *51*

D. Evidences for Pol II activity at tRNA genes *59*

E. Ser2-P affects the chromatin structure at tRNA genes *69*

F. The exonuclease Rrp6 targets pre-tRNAs for degradation 69

G. The SAGA connection 73

H. ADDENDUM: A global Pol II transcriptional defect in the S2A mutant ? 77

3. Conclusion 81

Part 2: Phosphorylation of Ser2 on the RNA polymerase II CTD is required for efficient trans-splicing of polycistronic transcripts in *C. elegans*. 85

1. Introduction 85

A. C. elegans development 85

B. Operons and trans-splicing in C. elegans. 87

C. CTD-Ser2 kinases in C. elegans 88

D. Objectives 89

2. Results & Discussion 89

A. An analog-sensitive allele of the Ser2 kinase Cdk-12. 89

B. Cdk-12 is required for post-L1 development 91

C. Cdk-12 inhibition impairs the expression of a subset of development genes. 91

D. CDK-12 inhibition does not globally impact the pre-mRNA splicing or 3' end formation. 97

E. Genes in operons are specifically affected by CDK-12 inhibition 99

F. Cdk-12 is required for efficient SL2 trans-splicing of genes in operons 99

G. Deficient SL2 trans-splicing leads to premature termination of Pol II transcription 103

H. Specific CDK-12-dependent induction of the SL2 trans-splicing upon recovery. 108

I. Other roles for CDK-12 activity? 108

3. Conclusions 111

MATERIAL & METHODS 113

1. Experimental procedures 113

2. Bioinformatics analysis 117

Appendix I 119

Appendix II 129

REFERENCES 139

The most exciting phrase to hear in science, the one that heralds new discoveries, is not *'Eureka!'*, but *'That's funny ...'*

ISAAC ASIMOV

ACKNOWLEDGMENTS | REMERCIEMENTS

Comme aurait pu le dire Otis dans un de ses monologues, une thèse, en résumé, c'est d'abord des rencontres. Et celles que j'ai eu la chance de faire durant ces quatre années m'ont permis, à elles seules, de surmonter les épreuves et défis propres à la thèse.

Tout d'abord, je me dois de remercier Damien. Avoir un chef qu'on estime, inspirant, *"someone to look up to"*, un exemple quoi… c'est inestimable. Toujours de bon conseil, attentif, positif, tu as été pour moi un véritable guide à travers ce sac de nœuds que peut être une thèse. Tu as su instiller en moi l'amour des belles questions de biologie, de celles qui vont au-delà des aspects purement descriptifs et mécanistiques, et ça, je le garderai toujours avec moi. Merci de m'avoir laissé la liberté d'explorer plusieurs chemins, malgré parfois des doutes légitimes (qui se sont d'ailleurs souvent vérifiés ☺). Merci aussi de m'avoir donné l'opportunité de m'attaquer à d'autres questions que celles de mon projet de thèse à travers des collaborations variées ; l'enrichissement personnel et professionnel que ça m'a apporté n'a pas de prix.

Puisqu'on parle des collaborations, merci à Antonin et Maxime d'avoir pris le risque d'accueillir un biologiste inexpérimenté, c'est-à-dire moi, en début de thèse, au sein de leur labo à l'institut Curie ! Maxime, tu as été d'une aide essentielle à la préparation des banques NGS, mais surtout, en tant qu'ex-belge, tu m'as permis de m'intégrer socialement et linguistiquement dans l'équipe. D'ailleurs, il faut que je te dise : même si parfois j'ai eu dur, grâce à toi, j'ai su tirer mon plan pour faire les banques. Aussi, je me souviendrai toujours de cette chouette soirée où on a fait la queue pour aller voir un certain humoriste dans Paris (encore un belge d'ailleurs), même que ça caillait et qu'il a draché après !

Antonin, rien ne t'obligeait à te comporter comme un mentor pour moi, et pourtant, tes conseils avisés et ton expérience m'ont réellement aidé à prendre certaines décisions importantes, alors merci du fond du cœur pour ça. Bon, rien ne t'obligeait non plus à me lancer vigoureusement ta serviette au visage au *3'end meeting* à Oxford quand je "draguais" malencontreusement une certaine PI en lui demandant de quel labo elle venait (et, selon la légende, si elle habitait encore chez ses parents), mais c'était de bonne guerre (si tu ne vois pas de quoi je parle, à mon avis Maxime s'en souvient encore). A quand la prochaine retraite à Nîmes ? C'était super !

Que serait la vie de tous les jours au labo sans ses collègues géniaux de GEMO ? Olivier, je ne sais même pas par où commencer pour parler de toi. Depuis notre rencontre à l'accueil des bleus bio il y a presque 10 ans, tu as pris une place spéciale dans mon cœur (non ce n'est pas une lettre d'amour, enfin j'essaie). Entre les Caras dans l'évier rempli d'eau froide (le frigo du pauvre) dans notre chambre des Chamois en Savoie, jusqu'aux ΔIPA et autres spéciales dans le frigo de la Beer Hour, en passant par notre tournée des

bars avec Constant à des fins documentaires[1], tu incarnes la joyeuse guindaille dans toute sa splendeur ! Mais au-delà de cet aspect purement festif, durant ces quatre années de thèse, j'ai aussi découvert un super collègue, dont l'intelligence, la motivation, l'abnégation, le don de soi et la persévérance ont su m'inspirer et m'inspirent encore. En plus, travailler à tes côté est toujours amusant et tu n'es jamais à court d'idée pour divertir tes chers collègues. Entre parties de mimes, rébus à gogo, charades à la chaine, jeu du CLAP-CLAP, et j'en passe, avec toi, il n'y a jamais vraiment de temps d'incubation ! Bon il t'es arrivé de t'amuser à mes dépends (par exemple, je ne te remercie PAS de m'avoir collé 1 236 133 trucs dans le dos), mais ce n'était jamais bien méchant et toujours dans la bonne humeur ☺ Alors mon petit Oli, Poppy, Popo, ou tout ce qu'on veut, merci d'être toi, et surtout ne change pas !

Constant (aussi connu sous le nom de Clément), toi aussi finalement on se connait depuis longtemps, même si on s'était un peu perdu de vue après le stage ~~en cara~~ en Savoie, avant de se retrouver avec plaisir en GEMO ! Constant, notre C-elegantiste pionnier qui nous a quittés trop tôt (c'est-à-dire qu'il a juste fini sa thèse deux ans avant Oli et moi), tu es aussi un artiste à la sensibilité brute (mi-ré-mi-ré-mi-si-ré-do-la) et au charmant accent de nos campagnes. Amoureux de la bière et de l'art (et aussi du lard) tu es le roi de la punch-line répétitive qu'on adore, et je pense encore tous les jours à tes fameux *"Coucou petite perruche", "Hey ! Pas plus de 12 nœuds !", "On va faire un tour ?", "Tu veux du saucisson ?"* Alors oui, je veux bien faire un tour avec bibiche et manger du saucisson avec toi quand tu veux !

Fanélie (Fanéliiiiiiie), la meilleure post-doc de GEMO, et peut-être bien de l'univers entier ! Ce que tu as apporté à l'équipe du haut de ta grande expérience (☺) n'a pas de prix : bricolages ingénieux, humour sale (oui, je rejoins Oli là-dessus), blagues pourries, rigueur scientifique (comment ça je mélange tout ?) … bref, c'est aussi en voyant les qualités et compétences de la brillante post-doc que tu es que j'ai eu l'envie de tenter le coup de mon côté. Merci pour tes conseils toujours avisés, nos discussions à ta pause de 17h qui se terminaient invariablement par un *"oublie pas d'prendre ton bus !"*, ta participation enthousiaste à toute activité sociale et j'en passe ! Puisses-tu trouver le bonheur dans la suite de ta carrière, qui, si elle doit prendre quelques virages, n'en sera pas moins pleine de succès et épanouissante, j'en suis sûr.

Ju', tu n'as pas eu peur de tendre la main au petit mémo que j'étais il y a 5 ans. Merci de m'avoir montré plein de trucs (scientifiques, hein), alors que rien ne t'y obligeais. Je m'excuse encore d'avoir contaminé tes transfos, je le ferai plus, promis ! (Il parait que j'ai appris à maniper depuis :p). On se voit bientôt au pays des caribous et du sirop d'érable !

Pol (aussi connu sous le nom de Philippe), mon maître, mon mentor, mon ami, si tu étais maitre jedi, je serais ton padawan. Si j'ai encore du mal avec certains aspects de

---

[1] Olivier était reporter pour le Guido à l'époque et avait pour mission d'écrire un guide sur les bars de Namur… Une aventure qui s'est terminée à coup de mi-ré-mi-ré-mi-si-ré-do-la (do-mi-si-lié) en chevauchant des poubelles communales (qu'on a bien sûr remises en place après – suite à l'intervention délicate d'un policier légèrement inquiet de notre utilisation d'un bien commun).

ton apprentissage (tes imitations de Jarjar Binks et tes *Blblblblblblbllbl* sont… inimitables), j'ai bien intégré ta devise : *« Il faut être l'ennemi de l'à peu près »,* et j'arrive à ne plus dire *« plus ou moins »* trop souvent, enfin… plus ou moins. Si j'en suis là aujourd'hui, c'est peut-être grâce à toi. Merci de m'avoir à peu près tout appris et d'avoir plus ou moins ouvert la voie d'un sujet de presque thèse, qui, s'il est assez difficile, n'en demeure pas moins relativement intéressant, enfin je pense.

Val', à défaut d'être un pilier de comptoir (d'ailleurs, ça change par rapport aux personnes de GEMO précédemment remerciées), tu es LE pilier du labo. Au fil des thèses qui s'enchaînent, on a pu assister à une escalade, bien justifiée, quant aux qualificatifs qui t'étaient adressés. Par ordre chronologique : "*Ma technicienne préférée*" – Pol ; *"Ma collègue de tong"* – Ju' ; *"La meilleure technicienne (Premier Agent Spécialisé Principal)"* – Constant ; *"La plus scintillante et pétillante étoile de la galaxie"* – Oli. Face à cette déferlante, je ne peux QUE surenchérir : Val, ton sourire rayonnant et ta pétillance suprême (de potiron) réchauffent les cœurs les plus frigides de Mars à Bételgeuse. Même si, contrairement à ce qu'en dit la légende, tu manges entre les repas, ta svelte silhouette est la preuve vivante que le temps n'a aucune emprise sur toi (sauf parfois quand le soleil brille dans la cuisine GEMO – HAAAAA UNE SORCIERE !!!). Sérieusement, jamais avare de conseils et de services, tu as été d'une aide précieuse pendant toute la durée de ma thèse et aussi de mon mémoire où tu as eu la patiente de me montrer ma toute première vraie manip ! Petite pensée aussi pour ton mari Oli, qui doit parfois bien rire quand tu lui racontes nos aventures quotidiennes au labo.

Malheureusement, malgré (ou grâce à) la qualité de toutes ces personnes qui ont contribué à mon apprentissage scientifiques et/ou bibitif, on ne peut pas rester élève toute sa vie. Quand est venu le temps pour moi de passer de l'autre côté du miroir et d'encadrer un étudiant sur mon projet, j'ai eu la chance de tomber sur Ysaline. Au premier abord, Ysaline, j'avoue, c'était mal barré, car malgré un clair potentiel, tu ne bois que des Mojitos et de l'eau. Heureusement, tu as su compenser ce petit défaut (et ta propension à faire des figures moches ☺) par de grandes qualités scientifiques : curiosité, indépendance, persévérance, esprit critique… J'ai vraiment eu de la chance de t'avoir comme étudiante, et tu peux être fière du chemin accompli sur ces 10 mois de mémoire. Je te souhaite bien du succès pour la suite et j'espère que la vie te sourira quelles que soient les voies que tu choisiras d'emprunter.

J'en profite pour remercier les autres mémorants du labo, anciens et actuels, car on aura quand même bien rigolé. Sarah et son légendaire Coca Cola (prononcer « Caucô Caulô »), Jérémy et ses histoires hallucinantes qui font passer un trip sous acide pour une promenade champêtre, Antoine et son imprimante, JB et son magnifique accent russe (toà bowarrrr veudekaaa), FX le roi de l'injection et de l'accent flamand débile (merci au fait de m'avoir fait découvrir Rothfuss), … Vous êtes tous une sacrée bande de zygoto, à croire que Damien filtre à l'entrée les gens trop normaux... Ne changez rien surtout, vous êtes parfaits comme ça !

Après ce « petit » tour d'horizon en GEMO, je me dois de remercier nos chers voisins d'URBM dont la convivialité est légendaire (ou comment l'aller-retour jusqu'à la machine à café peut prendre quasi une demi-journée).

Dans le désordre, merci à Séverin (et son remplaçant, Francesco) pour ses discussions (toujours passionnantes) en embuscade. Bruce, toi et moi, on est pareil sur certains points (« C'est qui s'vieux-là ? » – *regardent en direction du recteur de l'UNamur* – « Je sais pas. ») Bonne m*rd* pour la fin de ta rédac' ! Dr. Katy, pour avoir fait nos bacs ensemble, je peux te dire que tu as peu changé en 10 ans, et c'est tant mieux ! Tes histoires incroyables de malchance me font toujours beaucoup rire ; continue à prendre la vie avec le sourire. Jérôme, amoureux de football (de Cha' et de votre petite fille aussi je pense ^^) et surtout tellement drôle, je garde un souvenir ému de ton souper de thèse sur fond de Belgique-Panama ! Angy, je ne désespère pas de te montrer la beauté et la splendeur de le biologie eucaryote, peut-être au cours d'une autre discussion au coin du tableau noir près de la machine à café. Mat' (Mathilde/Marbite), tes talents de dessinatrices de certaines parties de l'anatomie humaine ne cessent de m'étonner. Rien à voir, mais il faudrait qu'un jour ton Shib' rencontre mon Shib', qui sait, ils seront peut-être potes. Mat' (Mathieu), ton humour pince sans rire n'a pas d'égal en ces lieux. Merci pour tout ce que tu fais pour nous, ce qui te rendscomplètement essentiel à la bonne marche du labo (je parle bien de ta propension à donner le coup d'envoi de la Beer Hour).

A propos de la Beer Hour justement, ce fut un plaisir de participer à l'organisation de cette fête hebdomadaire, une tradition qui tisse les liens sociaux au labo depuis 26 ans maintenant ! Parmi les habitués, je tiens à remercier Kriek (Gwen), Kwak (Agnes), Chimay bleue (Kévin), Leopold 7 (Pierre), Delirium Tremens (Pauline), Coca (Aurore & Dounz), Blanche de Namur (Bruce, mais c'est parce qu'il conduit) et tous ceux que je n'ai pas pu inclure dans la liste, car impossible de les associer à une bière particulière… ils les aiment toutes (Charles & co) ! Merci aussi à nos collègues d'URPhyM, notamment Gégé, Axe, Oli Svsk, Niam, François, JM et les autres pour leur accueil et leur convivialité.

Au-delà des collègues, je me dois de remercier le personnel du Gecko, pour leur délicieux plats Thaï, leurs sourires et leur gentillesse *"un plat de riz gratuit en plus, comme ça, manger gecko tous les jours"*. Un thésard bien nourri est un thésard heureux ! Merci aussi aux membres de l'harmonie de Fraire pour me permettre, jusqu'il y a peu (pour cause d'emploi du temps trop chargé), de me vider la tête le vendredi soir aux répétitions, une activité qui a été essentielle au maintien de mon équilibre mental au cours de ces quatre dernières années.

Merci aux amis et à la famille pour avoir toujours été là. Papa, tu as su me transmettre ta curiosité naturelle et ta passion pour les sciences. Maman, ton esprit critique m'a souvent servi d'exemple. Thomas, bro', nos discussions sur le chemin de l'harmonie me manquent, on ne se voit plus assez souvent ces temps-ci. Papy Luis, je te dédie cette thèse. Ta force tranquille, ton flegme et tes blagues en espagnol résonnent toujours en moi.

Merci à mes grands-parents de Dinant et ma tante Isa de m'avoir gentiment hébergé une partie de mon mémoire, et merci aussi à mes couz' Alex et Robin de m'avoir laissé l'ex-chambre de Céline pendant cette période. C'était très amusant de vivre avec vous tout ce temps, d'ailleurs je me souviens avec émotion des crabes royaux d'Isa, de mon gâteau-tartine, qui était vraiment le meilleur gâteau d'anniversaire du monde entier et de l'univers, et de nos apéros presque quotidiens :D

Virgil', si j'ai la force de me lever le matin depuis bientôt un an, c'est en partie pour pouvoir profiter de ton agréable compagnie sur la route de Dinant. Merci d'avoir accepté de me prendre en covoiturage, et d'être toujours là au rendez-vous *"bon pied-bon œil"* comme tu dis. Pourtant, dieu sait que c'est un peu difficile pour toi le matin, quand chaque seconde compte et qu'il faut éviter les pièges tels que les pots de shampooing presque vides ou les véhicules trop lents sur la route. Il n'empêche qu'en un an, on n'a jamais été en retard, ou en tous cas, jamais plus en retard que le train ! Je te souhaite beaucoup de succès avec ton projet asinien.

Il va sans dire que si j'en suis là aujourd'hui, c'est aussi et surtout grâce à ma petite famille. Morgane, tu es l'ancre de mon navire, et quand les flots de ma thèse m'emmenaient à la dérive, tu as toujours su tenir bon pour moi. Quand le soir venu, je vous retrouve toi et Yuk-Yuk, tous mes soucis s'évanouissent, tant vous savez toujours me faire rire avec vos bêtises et vos léchouilles (enfin, les léchouilles, ça concerne surtout mon chien Yuko). Si j'ai déménagé au fin fond de la Belgique pour toi, toi tu prévois de me suivre au fin fond du Canada. Alors merci pour ça, merci de m'avoir accepté comme je suis, avec mes folies et mes défauts, merci pour ton soutien inconditionnel, merci, merci, merci, je vous aime !

SHORT ENGLISH VERSION OF THE ACKNOLEDGMENTS

This thesis is the result of a highly collaborative work with inputs from Antonin Morillon and Maxime Wéry, who provided guidance and access to the high-throughput-sequencing facilities at the Curie Institute; from Yota Murakami and Takuya Kajitani, who nicely shared their ChIP-seq data of (phosphorylated) Pol II in fission yeast; from Clément Cassart and Fanélie Bauer who handled all the wet lab experiments for the *C. elegans* project; from Ysaline Lebrun who chose to work with me for her master thesis project and identified many *rpc25-flag* suppressors in *S. pombe*; from Valérie Migeot, who performed a nucleosome scanning experiment and might do more for this project before it ends; and from François Bachand, Ken-ichi Noma, Helena Idalgo, Dominique Helmlinger and Vincent Vanoosthuyse for nicely sharing some of their *S. pombe* strains.

A lot of credit must also go to Damien, who helped me navigate through my PhD with careful advices, genial ideas, positive thoughts and much more. Thank you for trusting me and giving me so much freedom.

I also have to acknowledge the EMBO and the F.N.R.S. for funding; the UNamur (especially the GeMo lab, Narilis, URPhyM and URBM) for providing the necessary infrastructure; and Alexandra Elbkyan for making science more open.

Finally, I would like to thank my friends, my colleagues, my family, my girlfriend and my dog for discussions and support.

Carlo

HOW TO READ THIS THESIS

i. Structure

This work is divided in two main parts that present two relatively independent research output produced during my thesis.

In the **first part**, I explored the causes and consequences of a surprising connection between Pol II phosphorylation and tRNA expression in fission yeast. Related to this part, the **Appendix 1** presents a published work assessing the conservation of the functions of the RSC chromatin remodeler (a candidate effector in the Pol II phosphorylation/tRNA expression relationship) in fission yeast and budding yeast (Yague-Sanz et al., 2017).

The **second part** of the thesis consists in a collaborative project in which we determined the role of the kinase CDK-12 in *C. elegans* development. With the exception of the first two figures that are required for the general understanding of the study, the presented research represents mostly my contribution to the project, that was limited to the generation and analysis of high throughput sequencing data. Related to this part, the **Appendix 2** presents a published bioinformatics pipeline that was specifically developed for the project in order to quantify special mRNA processing events (called *spliced leader trans-splicing* events) that we found were dependent on the activity of the CDK-12 kinase (Yague-Sanz and Hermand, 2018).

Preceding those two main parts of my thesis, a **general introduction** summarizes concepts important for the understanding and contextualization of both parts, while smaller introductions within the parts introduce more specific concepts.

ii. Gene/protein naming conventions

Gene names in *S. pombe* and *C. elegans* are always written in lower case italic. For instance: *lsk1*, *cdk-12.*

For *S. pombe* proteins, the names are written in straight lower case letters with the first letter capitalized (Lsk1). In contrasts, for *C. elegans* proteins, every letters are capitalized (CDK-12).

For the gene and protein names in *C. elegans*, a dash separates the alphabetic characters from the numeric characters (CDK-12, not CDK12).

ABBREVIATIONS

Term	Definition
ATP	adenosine tri-phosphate
a.k.a.	also known as
as	analog sensitive
CTD	C-terminal domain
C-terminal	carboxy-terminal
ChIP	chromatin immunoprecipitation
CPST	cleavage and polyadenylation specificity factor
CstF	cleavage stimulatory factor
DNA	deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
GMP	guanosine monophosphate
IGR	inter-genic region
lncRNA	long non-coding RNA
mRNA	messenger RNA
MAP	mitogen activated protein
m7G	N7-methyl guanosine
ncRNA	non-coding RNA
NDR	nucleosome depleted region
Ser2-P / Ser5-P	Phosphorylated serine 2/5 of the RNA polymerase II C-terminal domain
polyA	poly-adenosine
pre-tRNA (-mRNA)	precursor tRNA (mRNA)
RSC	remodel the structure of chromatin
RNA	ribonucleic acid
RP	ribosomal protein
rRNA	ribosomic RNA
RNAi	RNA interference
Pol	RNA polymerase
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
SL	splice leader
SAGA	Spt-Ada-Gcn5 acetyltransferase complex

General Introduction

1. Life & Transcription

i. What is life?

Biology is the science that study life, but *what is life ?* Physicists, chemists or geologists do not have issues in defining their own field, yet biologists struggle with this simple question, for which there is no clear-cut answer even today (Ferreira Ruiz and Umerez, 2018).

Even if some researchers actually argue that defining *life* is either impossible or pointless (Machery, 2012), I will attempt here to provide a short, partial definition of what I am studying, a.k.a life – but life in the context of the science pursued in my PhD thesis. As such, the following definition does not mean to be inclusive of every aspects of *life*:

Life: The ability to use internal information, in combination with energy, in order to carry various functions, such as regulating its internal environment, responding to stimuli, and, most importantly, propagating this internal information through reproduction or growth.

ii. The central dogma of molecular biology

According to this opinionated definition, one central aspect of life is the ability to use internal information. In life as it is now – but probably not in *life-as-it-was* (Alberts et al., 2002) –, biological internal information is stored in a 4-letters code, the DNA[2]. The use of that internal information, a process generally called *gene expression*, depends primarily on transcription, where RNA molecules (a similar 4-letters code) are synthetized according to the DNA template by DNA-dependent RNA polymerases (Pol). In turn, the RNA can carry a function *per se*, or serve as a template for protein synthesis from amino acids (a 20-letters code).

Even before the discovery of the first RNA polymerase (Stevens, 1960), the way biological information flows within the cells has been described by Francis Crick in an essay called 'On protein synthesis' (Crick, 1958). He observed that due to the nature of the biological codes (DNA, RNA, proteins), the biological information can in theory flow in any direction except from proteins to either DNA or RNA[3]. These insights were later

---

[2] Actually, there are exceptions to this rule, for instance in RNA viruses. Then again, whether the definition of life should include viruses or not is also controversial.

[3] This conclusion came from the observation that one needs at least 3 elements (later called codons) of a 4-letters code to unambiguously build a 20-letters code. However, as there are

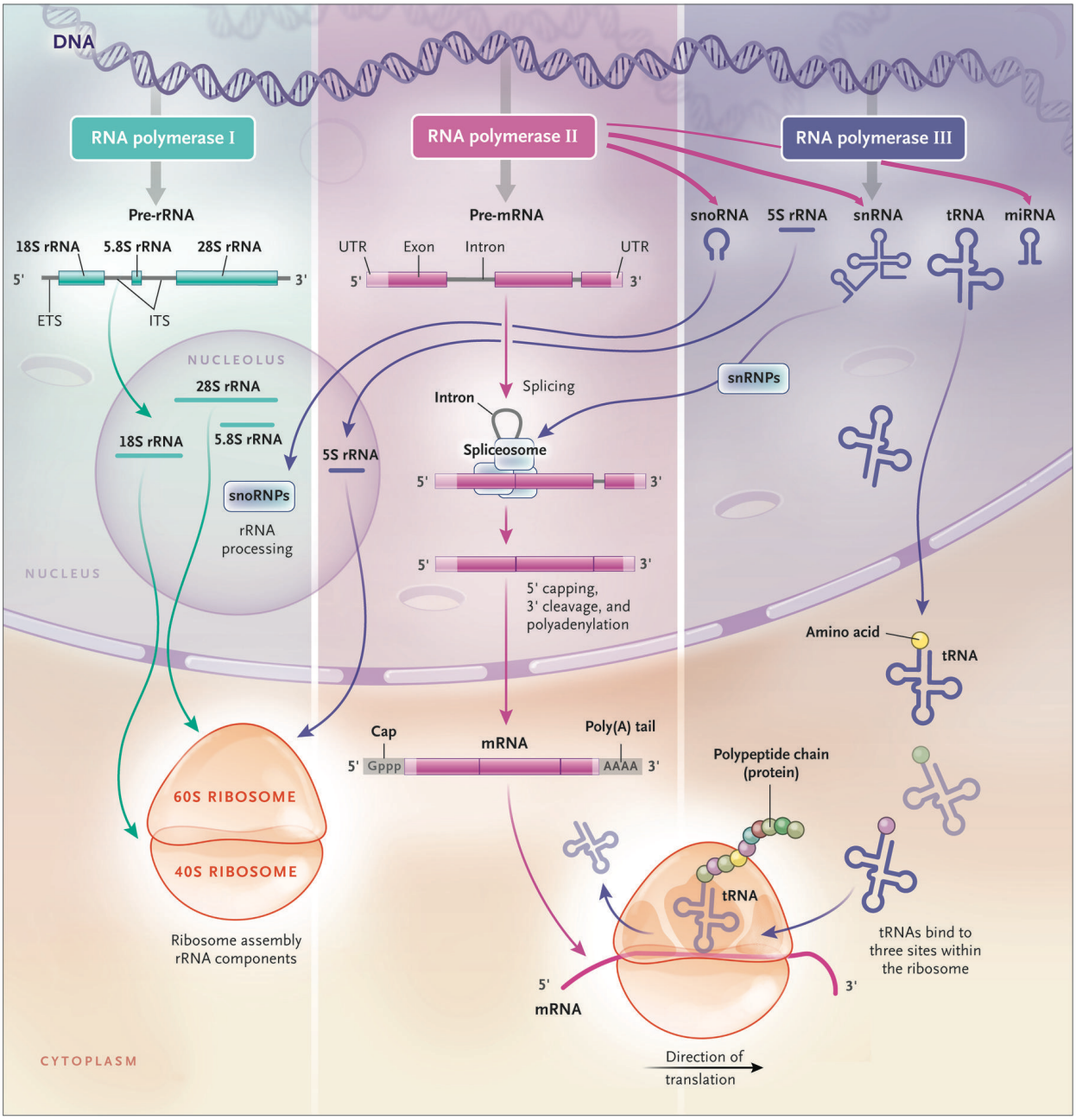


Figure 1: A world of RNAs

See main text for details. Adapted from (Lieberman et al., 2018)

ETS denotes external transcribed spacer, Gppp denotes guanosine triphosphate, ITS denotes internal transcribed spacer, and UTR denotes untranslated region.

referred to as the "the central dogma of molecular biology" and have been often oversimplified as an unidirectional flow of biological information (DNA --> RNA --> proteins), that is often true but for which there are countless exceptions. However, as initially stated, the central dogma still holds true.

In this work, we especially focused on how the biological transfer of information occurring during to first step of gene expression (DNA --> RNA) is regulated, both at a basal level and in response to external stimuli.

iii. The complexity of transcription

In eukaryotes – that is, living organism that have nuclei to compartmentalize their DNA within the cells – there are three main DNA-dependent RNA polymerases (Pol) called Pol I, Pol II and Pol III (Roeder and Rutter, 1969). These polymerases are evolutionary-related proteins complexes that specialized to transcribe specific sets of genes (**Figure 1**).

Pol I transcribes the main ribosomal RNA (rRNA) transcript, which encodes three of the four rRNAs (5.8S, 18S, and 28S) that compose the ribosomes, *i.e.*, the machinery of protein synthesis.

In parallel, Pol II transcribes "coding" messenger RNA precursors (pre-mRNAs) that, after maturation into messenger RNAs (mRNAs), will be used as template for protein synthesis. Pol II also transcribes a variety of non-coding RNAs (ncRNAs) such as small nucleolar RNAs (snoRNAs), important for rRNA processing into functional ribosome components; small nuclear RNAs that assemble with proteins to form the spliceosome, a complex involved in the maturation (splicing of the introns) of the pre-mRNAs; and a plethora of micro RNAs (miRNAs) and other ncRNAs (not shown on figure 1) that are mainly involved in the regulation of transcription.

Finally, Pol III mostly transcribes transfer RNAs (tRNAs) that allow the conversion of the 4-letters RNA code into the 20-letters protein code during protein synthesis. Pol III also transcribes other small ncRNAs such as the *snU6* snRNA or the *5S* rRNA.

Altogether, the intricated network of coding or functional RNAs show the complexity of the transfer of information that occurs within *life*.

---

64 possible combinations of groups of 3 elements from a 4-letters code, these combinations define the 20-letters in a redundant way (the code is degenerated). In consequence, it is impossible to come back to a 4-letters code (RNA or DNA) from the 20-letters code (proteins).

2. THE FISSION YEAST *S. POMBE*

i. Yeasts: powerful eukaryotic model organisms

Yeasts are part of the fungus kingdom, but in contrast with your typical white mushroom, they are primarily unicellular eukaryotes (although some species are able to form multicellular structures under specific conditions). Interestingly, single-celled yeasts appear to have evolved from multicellular ancestors, illustrating that evolution does not always favors more complex, "higher", organisms (James et al., 2006). Indeed, simpler features, like to be unicellular, can be evolutionary advantageous when it comes to multiply faster than competitors. For instance, the fission yeast *Schizosaccharomyces pombe* can complete its life cycle and give progeny within 2.5 hours, potentially multiplying by a thousand the number of cells in a population in little more than 24h.

Because of their fast growth rates, which make them convenient to culture and harvest, and also because of their ability to integrate foreign DNA into their genome through homologous recombination, which facilitates genetic manipulation (Fowler et al., 2014), yeasts have been used as eukaryotic model organisms for decades. Two species have received particular attention: the baker's (or budding) yeast *Saccharomyces cerevisiae* and the fission yeast S*chizosaccharomyces pombe*.

S. cerevisiae is a very interesting creature. Since at least 7000 years – 13 000 years according to a recent study (Liu et al., 2018) –, humans have taken advantage of its intrinsic ability to ferment in hypoxic conditions to bake bread and to brew a variety of alcoholic beverages (Mortimer, 2000). Having evolved in contact with humanity for thousands of years, the well named baker's yeast is now considered a domesticated species, although some wild isolates still remain (Greig and Leu, 2009). Given this long history (and the relative smallness/simplicity of its genome) it is only natural that *S. cerevisiae* was the first eukaryotic species whose genome have been fully sequenced, entering the era of genomics 5 years before the publication of the human genome (Goffeau et al., 1996; Lander et al., 2001; Venter et al., 2001).

In contrast, there is little historical application[4] of the fission yeast S*chizosaccharomyces pombe* and as such, the interest for this yeast as a model organism is rather recent and only started in the second half of the 20th century (Schlake and Gutz, 1993). Nonetheless, this organism has been rapidly considered as a central model organism in molecular genetics and provides advantageous features complementary to those of the budding yeast – reviewed in (Hoffman et al., 2015) and briefly summarized hereafter.

---

[4] One notable exception is that *S. pombe* was used in African countries for its alcoholic fermentation. Accordingly, "pombe" means "booze" in Swahili. Note that some *S. pombe* researchers have experimented brewing with it and all agree to say that it tastes awfully [discussion on the pombelist]. Recently, more successful application of *S. pombe* in rum and wine making have been developed, but remain at a very small scale (Volschenk et al., 2003).

ii. Why is fission yeast a great model organism?

S. pombe is an ancient yeast whose roots go back, according to phylogeny, to the early evolution of fungi (Sipiczki, 2000). In consequence, the evolutionary distance between *S. pombe* and *S. cerevisiae* is on the same order as the distance between either yeasts and mammals. However, *S. pombe* appears to have evolved at a slower rate than *S. cerevisiae* and retains more characteristics from the last common eukaryotic ancestor, which are often shared with mammals but absent in *S. cerevisiae*. For instance the complete set of proteins required for RNA interference (Dicer, Argonaute and the RNA-dependent RNA polymerase) are conserved *in S. pombe* and have been well studied, leading to major discoveries in the field – reviewed in (Volpe and Martienssen, 2011). In addition *S. pombe* has large centromeres, structurally similar to that of mammals, on which RNAi-dependent heterochromatin is deposited (Volpe et al., 2002). All those features established the fission yeast *S. pombe* as one of the major models for the study of molecular and cellular biology in eukaryotes.

iii. The fission yeast life cycle

In rich conditions, when nutrient availability is high, the fission yeast *S. pombe* grows exponentially. As its lay name suggests, the fission yeast divides by medial fission. Partly because of these symmetric division that allow researchers to assess the age of a cell – and cell-cycle defect – just by looking at its size, study in fission yeast lead to major breakthroughs in the area of cell-cycle control. For a personal recollection from sir Paul Nurse on how his research on the regulators of the cell-cycle in fission yeast led him to receive the Nobel prize in 2001, see (Nurse, 2017).

The exponential growth in *S. pombe* corresponds to a succession of mitosis (M phase), a short growth phase (G1), DNA replication (S phase), and a longer growth phase (G2) (Forsburg and Nurse, 1991). It takes about 2h30 for this vegetative cell cycle to complete in optimal conditions.

However, upon nutrient starvation, especially nitrogen starvation, cells stop their mitotic cycle at the G1 stage and can undergo sexual differentiation. First, a transient diploid (the fission yeast is haploid during its vegetative cell cycle) is generated by the conjugation of two cells of opposed mating types called *h+* and *h-*. The diploid cell (also called zygote) will then proceed through meiosis I and II to form an ascus containing four spores. The spores are metabolically arrested haploid cells that are resistant to a variety of stresses and can survive for months in harsh conditions. However, when the conditions becomes favorable again, the spores will re-enter the mitotic cell cycle (Egel, 1973) (**figure 2**).

The irreversible switch from proliferation (mitotic cell cycle) to sexual differentiation can be considered as the most important cell-fate decision in the life of *S. pombe*. As such, it is tightly regulated in response to environmental cues by a variety of signaling pathways – reviewed in (Anandhakumar et al., 2013) – that converge to control the expression of Ste11, the master regulator of sexual differentiation.

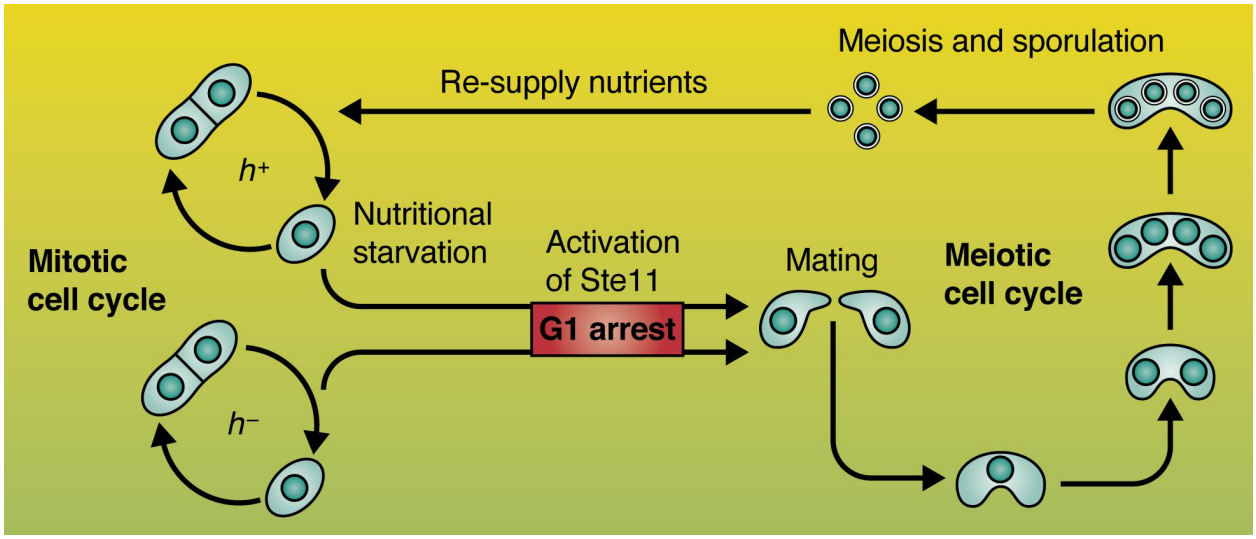


Figure 2: *Schizosaccharomyces pombe* life cycle

Adapted from (Otsubo et al., 2012). See main text for details

3. THE RNA POL II C-TERMINAL DOMAIN

A. Complexity of the Pol II C-terminal domain (CTD)

Among all RNA polymerases, Pol II received the most attention, as it transcribes mRNA with coding potential. The Pol II complex is composed of 12 subunits (named Rpb1 to Rpb12, by decreasing order of molecular weight). The largest subunit, Rpb1, includes a unique and flexible domain at the carboxy-terminal (C-terminal) end of the protein. This intrinsically disordered extension, called the C-terminal domain (CTD) is composed of sequential repeats of a consensus $Y_1S_2P_3T_4S_5P_6S_7$ heptapeptide[5] (Corden et al., 1985). While the consensus sequence is conserved among all eukaryotes, the number of repeats and hence, the length of the CTD, varies and tends to be correlated with the apparent complexity of an organism (Allison et al., 1988). For instance, the number of repeats can range from 26 repeats in budding yeast, 29 in fission yeast, 42 in the nematode *C. elegans* and 52 in mammals (Corden, 1990).

Work in *Drosophila melanogaster* showed that, although the CTD is essential for viability *in vivo*, it is dispensable *in vitro* for accurate Pol II transcription (Zehring et al., 1988). In budding yeast, the CTD can be shortened to 13 heptad repeats and still be perfectly viable for the cell, but further deletion down to below 10 repeats is lethal, indicating that the heptad repeats are partially redundant for the essential role of the CTD (Nonet et al., 1987). In addition, the CTD can be physically transferred from Rpb1 to the Rpb4 or Rpb6 Pol II subunits without affecting cell viability (Suh et al., 2013). Moreover, replacement of the highly degenerated CTD of drosophila with the human CTD is also perfectly viable (Portz et al., 2017). Suggesting that the minimal functional unit of the CTD consists of in two consecutive heptads, genetic analysis in yeast revealed that insertion of additional amino acids between single heptads is lethal, while insertion between two consecutive heptads is tolerated (Stiller and Cook, 2004).

Altogether, these aforementioned studies highlight that while the CTD is an essential domain, it is not required for Pol II transcription *per se*. Instead, it appears that the CTD function as a recruitment platform for multiple factors involved in co-transcriptional processes (detailed in the next chapters). The specificity of the recruitment of those diverse factors has been attributed to specific configurations adopted by the – otherwise disordered – CTD due to the combination of post-translational modifications and conformation changes within the CTD, termed the "CTD code" (Buratowski, 2003) .

Indeed, the CTD is extensively post-translationally modified, potentially on each residue of the consensus heptads. In abundance, the most important modifications are the phosphorylation of the serines S_2 (Ser2-P) and S_5 (Ser5-P) of the consensus repeats (Heidemann et al., 2013; Suh et al., 2016), but the tyrosine (Y_1) (Baskaran et al., 1993), the threonine (T^4) (Hsin et al., 2011), and the serine S_7 (Egloff et al., 2007) can also be

[5] Y= tyrosine; S= serine; P= proline; T= threonine.

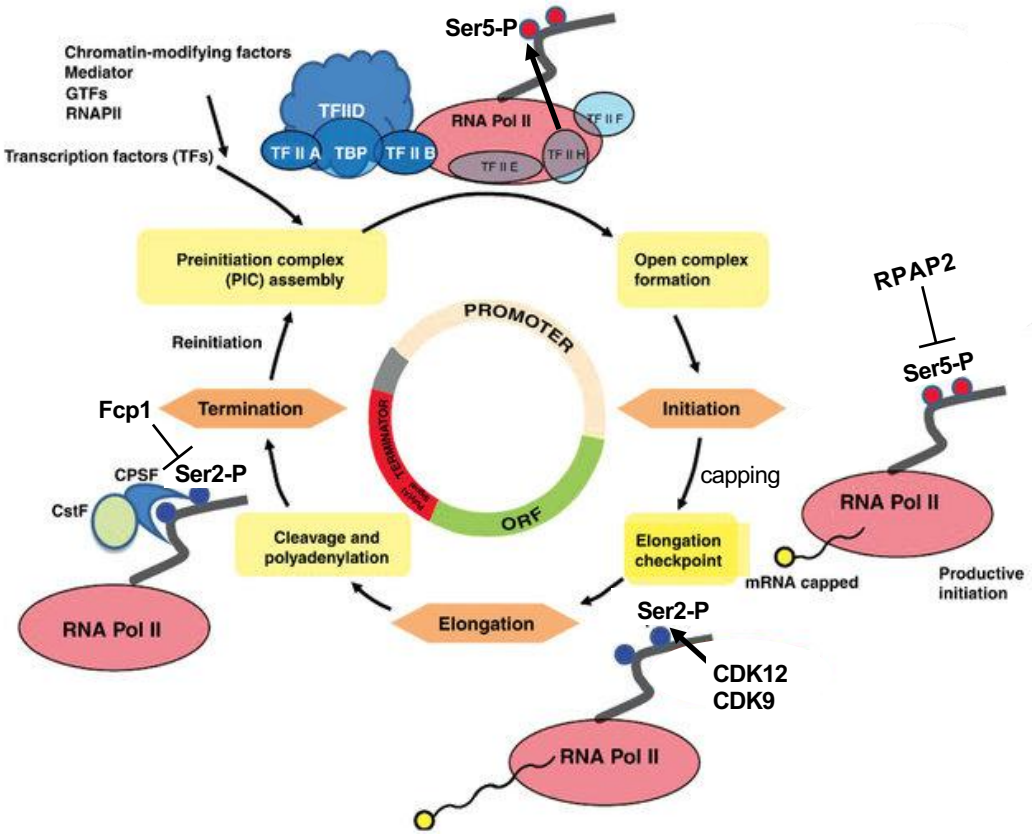


Figure 3: Pol II transcription cycle

Adapted from (Hong et al., 2016). See main text for details

phosphorylated. Other modifications include S_2, S_5 and T_4 glycosylation (Lu et al., 2016) and the methylation, acetylation and ubiquitination of arginine residues in the non-consensus repeats (Schröder et al., 2013; Voss et al., 2015). In addition, the two prolines (P_3 and P_6) are subject to isomerization and can be found in either *cis* or *trans* conformations (Hanes, 2014).

Deciphering the functional importance behind the – potentially very complex – CTD code has been the focus of numerous studies in the past two decades. However, it has been shown recently by using mass spectrometry (instead of antibody-based assays) that the complexity of the CTD code might have been overestimated in the past, as the phosphorylation of Ser2 and Ser5 were found a 100-folds more abundant than other phosphorylated residue (Suh et al., 2016). While the absolute abundance of a modification is not a measure of its biological relevance (a modification could be present only transiently during the transcription cycle, yet be biologically important), we will, for the sake of brevity, mainly focus on the roles of the more abundant Ser2-P and Ser5-P for the next chapters of this introduction.

B. Pol II CTD coordinates mRNA maturation with transcription

i. The transcription cycle

Chromatin immunoprecipitation experiments using antibodies specific to Ser2-P and Ser5-P revealed that these modifications are brought in an ordered fashion, sequentially on the transcription units and are correlated with transcription cycle (Komarnitsky et al., 2000).

Before initiation, unphosphorylated Pol II is recruited to the promoter. Then, CDK7 (Kin28 in buding yeast and Mcs6 in fission yeast), within the TFIIH complex, phosphorylates Ser5 upon transcriptional initiation (Cismowski et al., 1995). Ser5-P peaks to its maximum level 20 to 120 nt downstream the initiation site (Mayer et al., 2010), but is then progressively removed by the phosphatases RPAP2 (Rtr1 in yeast) in the vicinity of the site of transcription initiation (Mosley et al., 2009). Dephosphorylation of Ser5-P continues toward the 3' end of the transcription unit by the Ssu72 phosphatase (Krishnamurthy et al., 2004).

In parallel with Ser5-P dephosphorylation, Ser2 is phosphorylated during transcriptional elongation by its major kinase CDK12 (Ctk1 in budding yeast, Lsk1 in fission yeast) (Cho et al., 2001; Coudreuse et al., 2010) and by its minor kinase CDK9 (Bur1 in budding yeast, cdk9 in fission yeast) (Qiu et al., 2009). Ser2-P peaks to its maximum level toward the 3'-end of the transcription unit, and is then removed by the Fcp1 phosphatase 100 to 200 nt downstream the polyadenylation site (Cho et al., 2001,Mayer, 2010 #100). Finally, the fully dephosphorylated Pol II terminates and can reinitiate a new cycle of transcription (**Figure 3**).

To sum up, the interplay of multiple CTD kinases and phosphatases successively recruited during the cycle of transcription contributes to create a uniform transition in the phosphorylation marks on Pol II CTD (Mayer et al., 2010). This differential CTD phosphorylation between the 5' and 3' end regions of the transcription unit allows the

differential recruitment of factors required at specific times of the transcription cycle as exemplified below.

ii. Capping

In eukaryotes, the nascent pre-mRNAs exiting the RNA polymerase needs to be protected rapidly to avoid degradation by exonucleases. To provide exonuclease protection, a cap of N7-methyl guanosine (m7G) is added on the 5'-end of the nascent transcripts in a 3-steps process. (1) First, a RNA triphosphatase hydrolyses the triphosphate of the first transcribed nucleotide into a diphosphate. (2) Then, a guanylyltransferase transfers a guanosine monophosphate (GMP) on the mRNA 5'-diphosphate end. (3) Finally a methyltransferase methylates the GMP on the N7 position (Mizumoto and Kaziro, 1987).

The two first steps of cap formation are catalyzed by the same enzyme in mammals (Mce1), while in yeast, two different enzymes are required (Pct1 and Ceg1 respectively). In mammals, Ser5-P connects the capping with transcriptional initiation through the recruitment of Mce1 (Ho et al., 1998). Similarly, in yeast, both Pct1 and Ceg1 are recruited by Ser5-P during the initiation (Proudfoot et al., 2002). This early recruitment of the capping enzyme(s) to the nascent transcripts allows their timely capping, an essential process to protect the transcripts from degradation. Actually, coordination of the capping with transcription is the only essential function of Ser5-P. This was elegantly shown by the covalent tethering of the mammalian capping enzyme Mce1 to the very end of the CTD, a tethering that bypasses the requirement of Ser5-P for viability in *S. pombe* (Schwer and Shuman, 2011).

iii. Splicing

In eukaryotes, many genes contain introns, i.e, stretches of nucleotides transcribes and present in the pre-mRNA that are removed by a process called "splicing" during mRNA maturation. Alternative splicing can generate diversity in the resulting mature mRNA (Pan et al., 2008), and the splicing process in general can serve as a regulator of gene expression (Li et al., 2017; Shkreta and Chabot, 2015). In mammals, almost all protein-coding genes have multiple large introns. The situation is different in *S. pombe* and *S. cerevisiae*, where the introns are much smaller and only 40% and 8% of the protein-coding genes (respectively) have introns (Hoffman et al., 2015).

The splicing of pre-mRNA is catalyzed by a huge ribonucleoproteic complex called the spliceosome. Although the splicing can occur independently than transcription *in vitro*, transcription is required for efficient splicing *in vivo* (David and Manley, 2011). It has been proposed that the CTD acts as a landing platform for the spliceosome (Greenleaf, 1993). However, the requirement of the phosphorylation of the CTD for splicing is controversial.

Supporting the idea of a CTD phosphorylation-dependent coordination of the splicing process, it has recently been shown that the splicing occurs during transcription, as soon as the 3' splice-site is being transcribed (Oesterreich et al., 2016). Moreover, the presence of a hyper-phosphorylated CTD increases splicing kinetic (Hirose et al., 1999).

In addition, physical interactions between the CTD and components of the spliceosome have been described (David et al., 2011; Morris and Greenleaf, 2000) and in human cells, a S2A mutant (where every Ser2 is mutated to alanine and cannot be phosphorylated) impairs the recruitment of the spliceosome (Gu et al., 2013).

However, arguing against a role (or at least, a conserved role) of Ser2-P in coordinating splicing, no splicing defect was detected in a genome-wide study using the S2A mutant in fission yeast (Inada et al., 2016). In addition, a recent studies that mapped nascent RNAs associated with phosphorylated Pol II in mammals found that splicing intermediates (and the spliceosome) were mostly associated with Ser5-P, suggesting that Ser5-P is the phosphorylated form the CTD responsible for the coordination of splicing in mammal (Nojima et al., 2015; Nojima et al., 2018). As elongating polymerases are phosphorylated on Ser2, it was suggested this connection between Ser5-P and splicing implies a transient shift of the phosphorylation pattern of the Pol II CTD as the splice sites are transcribed (Custódio and Carmo-Fonseca, 2016). However, the mechanism behind such shift of phosphorylation is currently unknown, promising further years of exciting research in decorticating how Pol II CTD coordinates splicing with transcription.

iv. Cleavage and poly-adenylation

The last step of pre-mRNA maturation is the 3'-end processing that consists in (1) the cleavage of the pre-mRNA and (2) the addition of a poly-adenosine (polyA) tail. Again, coordination with transcription is important because it is the cleavage that ultimately allows the polymerase to terminate according to the torpedo model of transcriptional termination (West et al., 2004) – although there are multiple and partially redundant fail-safe mechanisms of Pol II termination (Proudfoot, 2016). In addition, as the polyA tail protects the free 3' end of the pre-mRNA against degradation by 3'-5' exonucleases, polyadenylation must occur in a timely manner.

Assembly of the pre-mRNA 3' processing complex is initiated by its binding to conserved consensus sequences on the RNA molecule. First, the cleavage and polyadenylation specificity factor (CPSF) recognizes the A(A/U)UAAA polyadenylation signal (PAS) on the pre-mRNA. Then, both a uridine-rich downstream element (U-rich DSE) and the CPSF, participate in the recruitment of the cleavage stimulatory factor (CstF). Finally, the RNA is cleaved by an endonuclease subunit within CPSF and polyadenylated by an ATP-dependent polyA-polymerase – reviewed in (Richard and Manley, 2009).

Several subunits of the 3' processing complex, such as Clp1, Rna14, Tex1, Rtt103, CstF50 and Pcf11 are known to interact with Ser2-P (Harlen et al., 2016, Lunde, 2010 #101; Meinhart and Cramer, 2004). In addition, CTD truncation causes a 3'-end processing defects (McCracken et al., 1997). As the recruitment of the 3' processing complex is primarily assured by the *cis* sequences on the pre-mRNA, the involvement of Ser2-P could be a way to assure robust, perfectly timed, coupling of transcription with 3'-end processing and termination.

C. Pol II CTD coordinates chromatin remodeling with transcription

i. What is chromatin?

Nuclear DNA from one typical human cell, once unraveled, can be stretched up to two meters (McGraw-Hill, 2012). Therefore, there is a strong need for the cell to package DNA into a compact structure called chromatin. This compaction also protects from DNA damage and has important implication in controlling gene expression.

The smallest functional unit of chromatin is the nucleosome. Nucleosomes are 8-subunits complexes that wrap DNA on approximately 147 bp. They are composed of two heterodimers of histone H2A and H2B sandwiching a tetramer of histone H3 and H4 (Luger et al., 1997). Non-canonical histones also exist, such as the histone H2AZ and the histone CENPA, a H3 variant only present on centromeres (Folco et al., 2008) .

Chromatin has two distinguishable states: the heterochromatin (densely compacted) and the euchromatin (less compacted) (van Steensel, 2011). In contrast with euchromatin, DNA in the heterochromatin state is generally considered silent as the compact structure is not accessible to the transcriptional machinery. In addition, a complex range of more subtle chromatin states can be defined according to the so called "histone code": the combination of various post-translational modification (acetylation, methylation, phosphorylation, ubiquitination, …) on the N-terminal tails of the histones (Jenuwein and Allis, 2001).

ii. Determinants of nucleosome positioning

Genome wide studies of nucleosome positioning revealed that one nucleosome spans on average every 200 bp. Strikingly, the distribution of the nucleosomes within the genome is not random as there are clear preferred positions. In particular, nucleosomes directly downstream transcription starting sites (TSS) are strongly positioned while regions upstream TSS or downstream transcription end sites (TES) are enriched in nucleosome depleted regions (NDR) (Fan et al., 2010; Yague-Sanz et al., 2017).

This non-random distribution of nucleosomes within the genome is determined by a multiplicity of factors. First, the intrinsic properties of the underlying DNA sequence paint a landscape of nucleosome positioning preferences (Kaplan et al., 2009). On top of this landscape, the binding of transcription factor on the DNA and/or the action of ATP-dependent chromatin remodelers (such as the RSC complex) creates NDR (Yague-Sanz et al., 2017). From there, nucleosome positioning spreads from the NDR according to the statistical positioning theory (Mavrich et al., 2008): the simple presence of a barrier (NDR) restricts the positions a nucleosome can occupy in adjacent locations. Then, the next nucleosome is also statistically restricted by the previous one, and so on. As the distance with the closest barrier increases, the precision of the positioning decays. Among the positions allowed by the statistical positioning, preference would be given to those with favorable intrinsic DNA properties (Johnson, 2010).

iii. Chromatin affecting transcription

The chromatin state has a considerable influence on transcription. Besides heterochromatin regions, where the DNA is so densely compacted that transcription almost never occurs, promoter recognition by Pol II transcription factors and subsequent PIC assembly depends on DNA accessibility (Tirosh and Barkai, 2008). Moreover, Pol II has to cope with the transcriptional barriers imposed by nucleosomes to transcribe efficiently, and it was shown that Pol II actually pauses at each major histone-DNA contact sites (Kujirai et al., 2018).

In that context, the histone code regulates transcription either by recruiting other regulatory proteins, or directly by altering the affinity between DNA and nucleosomes. Typically, histone acetylation decreases the affinity between DNA and nucleosomes by neutralizing the positively charged histones. In contrast, histone methylation tends to favor nucleosome compaction by competing with acetylation for the same residue or by recruiting histone deacetylases, making it harder for the RNA polymerase to displace nucleosomes during elongation (Belch et al., 2010). Intriguingly, histone acetylation and histone methylation are both correlated with active transcription (Berger, 2002). An explanation for this apparent paradox will be proposed in the following paragraphs.

iv. Transcription affecting chromatin

As most co-transcriptional processes, chromatin remodeling and histone modifications are connected to Pol II CTD phosphorylation.

Shortly after Pol II initiation, the histone deacetylases Gcn5 within the SAGA complex and Esa1 within the NuA4 complex are recruited via Ser5-P to relax chromatin compaction and ensure productive elongation (Govind et al., 2007; Ng et al., 2003).

During transcriptional elongation, the RNA polymerases displace nucleosomes in their wake as they proceed through the genes. However, the displaced nucleosomes need to be replaced behind the polymerases in order to prevent spurious transcription initiation within the body of the gene (Workman, 2006) in a process dependent on Pol II CTD phosphorylation. At the early elongation phase of Pol II transcription, Ser5-P leads to the recruitment of the Set1 histone methyltransferase (part of the COMPASS complex) responsible of the methylation of the lysine 4 on the histone H3 (H3K4me) to the 5' ends of transcribed genes (Ng et al., 2003). Similarly, as the elongation proceeds, Ser2-P recruits the Set2 histone methyltransferase that methylates the lysine 36 on the histone H3 (H3K36me) toward the 3' end of the gene (Kizer et al., 2005). H3K4me3 and H3K36me then respectively recruit the histone deacetylases Set3 and Rpd3 (Keogh et al., 2005; Kim and Buratowski, 2009), which deacetylates the histone displaced by Pol II, favorizing chromatin recompaction in the polymerase wake.

In short, a common role of Ser5-P and Ser2-P in regards to histone modifications would be to restrict the spread of chromatin acetylation to promoters and to help reposition nucleosomes in the wake of the RNA polymerase. Indeed, it was reported that in the absence of these mechanisms, long nucleosome depleted regions lead to cryptic initiation of transcription within the genes, resulting in the transcription of aberrant RNAs (Guillemette et al., 2011; Quan and Hartzog, 2010).

D. Gene- or condition-specific roles for CTD modifications

i. Gene-specific requirements of Ser2-P

Challenging the classical view of CTD phosphorylation, stating that a uniform transition occurs from Ser5-P to Ser2-P as the Pol II proceed spatially and temporally through transcriptional elongation, works from our group revealed that a subset of genes exhibit promoter-proximal Ser2-P in fission yeast. This early Ser2-P correlates with the recruitment of its kinase Lsk1 in the promoter region of the concerned genes, including the master regulator of transcription *ste11*. Importantly, loss of Ser2-P either by deleting Lsk1 or in the S2A mutant (where every Ser2 within the CTD are mutated into alanines) is perfectly viable, but causes a penetrant sterility that was attributed to a defect in the expression of *ste11,* the gene encoding the master regulator of sexual differentiation *(Coudreuse et al., 2010)*.

Mechanistic insights into this gene-specific regulation revealed that promoter-proximal Ser2-P relieves *ste11* from the Ser5-P-dependent transcriptional repression imposed by Set1 H3K4 methylation and the subsequent recruitment of histone deacetylases (Materne et al., 2015). Downstream this histone modification cascade, the RSC complex was later shown to be important for *ste11* induction in creating a nucleosome depleted region at *ste11* promoter. In contrast, H3K4 methylation, and also the ubiquitination of H2B, initiates the deacetylation process, which decreases chromatin remodeling by the RSC complex (Materne et al., 2016).

Therefore, besides challenging the broadly accepted model where Ser2-P is uniformly distributed toward the 3'-end of the genes, these studies point toward a gene-specific requirement of Ser2-P for gene expression. Similarly, gene-specific expression defects due to the S2A mutation on a truncated CTD were also reported by another group, also working in *S. pombe* (Schwer et al., 2014). In *Drosophila melanogaster*, depletion of the Ser2 kinase Cdk12 specifically affects the expression of stress-activated genes (Li et al., 2016) and causes ectopic heterochromatin formation on long genes involved in neuron function (Pan et al., 2015). Finally, in mammalian cell cultures, the Ser2 kinase CDK12 depletion specifically affects genes implicated in the DNA damage response (Blazek et al., 2011b; Ekumi et al., 2015; Liang et al., 2015).

ii. Condition-specific induction of CTD phosphorylation

Despite the basal functions of CTD phosphorylation during the transcription cycle, our work and others revealed that CTD phosphorylation is also regulated in response to external stimuli. Indeed, CTD Ser-2 phosphorylation is globally increased in response to nitrogen starvation (Sukegawa et al., 2011). This stress response is mediated by the Sty1 MAP kinase that directly phosphorylates the Ser2 kinase Lsk1 on its N-terminal domain. The phosphorylation of Lsk1 boosts its activity and mediates its recruitment to the *ste11* promoter, allowing the early Ser2-P to derepress *ste11* expression as described in the previous paragraph (Materne et al., 2015; Materne et al., 2016).

Similar modulation of CTD phosphorylation in response to external or internal cues have been described in other systems. For instances, in *S. cerevisiae*, nutritional stress and heat shock also result in higher levels of Ser2-P (Åkerfelt et al., 2010). In *Arabidopsis thaliana*, the activity of CTD kinases and phosphatases are modulated in

response pathogen perception (Li et al., 2014) and in human T-cells, HIV infection hijack the cellular CTD kinases and phosphatases to stimulate viral gene expression (Chen et al., 2014; Mbonye et al., 2013).

4. General objectives of this study

Following up on the gene-specific and condition-specific requirements highlighted by our work and others, the first part of this thesis builds on the surprising observation that the Pol III-transcribed tRNA genes are affected by the loss of Ser2-P at the chromatin level (unpublished observation from the data analyzed in Materne et al., 2015). In this work, we explore how the connection between Pol II CTD phosphorylation and Pol III transcription might provide a new mean of coordinating the expression of both polymerases, possibly in response to environmental cues.

In the second part of this thesis, we develop a rather different aspect as we explore the role of the Ser2-P kinase CDK-12 in the development of a multicellular nematode, *C. elegans*. Again, we uncovered gene-specific requirements for CDK-12 and hints that CDK-12 might be, as for its homolog in *S. pombe*, regulated in response to environmental and developmental cues.

PART 1: PHOSPHORYLATION OF SER2 ON THE RNA POLYMERASE II CTD OPPOSES TRNA EXPRESSION.

Contributions:

Insight on and access to the high-throughput sequencing technologies:	Maxime Wéry & Antonin Morillon
ChIP-seq and RNA-seq libraries generation and analysis:	Carlo Yague-Sanz
tRNA northern blots, ChIP-qPCR and growth-assay experiments:	Carlo Yague-Sanz
Rpc-25 flag mutant characterization and identification of suppressors:	Ysaline Lebrun

1. INTRODUCTION

In this part of the thesis, we explore a functional and genetic interactions between Ser2-P and tRNA expression and propose the existence of a new regulatory layer that controls Pol III transcription. While a summary of the current knowledge on Pol II and Ser2-P was covered in the general introduction, we present hereafter a more focused overview of tRNA metabolism. The topics covered include the mechanisms of Pol III transcription, its regulation and the maturation process for tRNAs (summarized in **figure 0**). Finally, hints from the literature that Pol II might constitute a Pol III transcription factor will be introduced.

A. The Pol III transcriptional machinery.

From the three eukaryotic RNA polymerases (Pol), Pol III is the biggest with 17 subunits. Despite that 10 of these subunits are unique to Pol III, the general architecture of the complex is conserved with Pol I and Pol II. Among the unique subunits, some have clear homologs within Pol II and/or Pol I, and other form subcomplexes which are structurally and functionally similar to general Pol II transcription factors (Vannini and Cramer, 2012). For instance the Pol III subunits Rpc37 and Rpc53 compose a TFIIH-like complex that facilitate the formation of an open promoter complex (Kassavetis et al., 2010). Similarly, the function of the TFIIE complex that recruits and stimulates TFIIH is recapitulated in the Rpc82/Rpc34/Rpc31 Pol III subcomplex (Geiger et al., 2010). Thus, the extra subunits within Pol III do not define additional features of the complex, but instead could be assimilated to a permanent recruitment of general transcription factors to Pol III to accommodate promoter specificity and transcription processivity (Carter and Drouin, 2010).

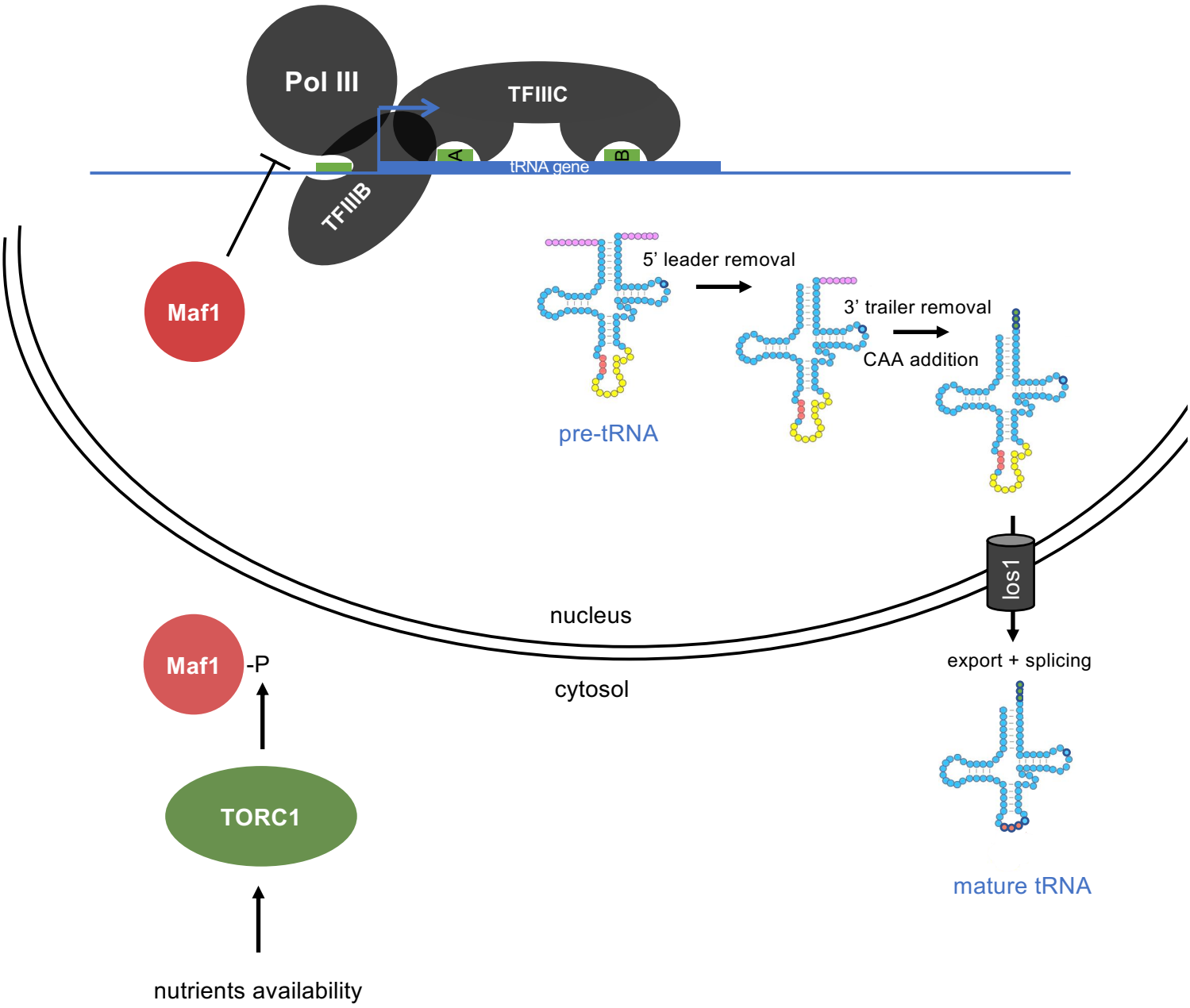


Figure 0: tRNA transcription, maturation and Maf1-dependent regulation

tRNA transcription: The Pol III transcription factor TFIIIC recognizes and binds the A- and B-boxes internal to the tRNA gene. Then TFIIIB is recruited by TFIIIC and the TATA box upstream of the tRNA gene. Finally, TFIIIB is able to recruit Pol III for transcriptional initiation/reinitiation.

tRNA maturation: The full length pre-tRNA is processed into mature tRNA by the removal of the 5' leader, the 3'-end processing (removal of the 3' trailer and CAA addition) and, if present, the splicing of the intron. The splicing occurs in the cytosol, at the outer mitochondrial membrane. Therefore, the pre-tRNA needs to be exported first.

Maf1-dependent regulation of Pol III transcription: When nutrients are available, Maf1 is phosphorylated by TORC1 and other kinases. In consequence, it localizes in the cytoplasm, preventing its repressive activity. However, in response to unfavorable conditions, Maf1 is dephosphorylated and is allowed to enter the nucleus, where it represses Pol III transcription

There are only two general transcription factors for Pol III: TFIIIB and TFIIIC. A third transcription factor, TFIIIA, is specifically required for the transcription of the 5S rRNAs and acts as an adapter for the binding of TFIIIC.

TFIIIC can be regarded as a pioneer transcription factor that, through strong interactions with specific intragenic promoter elements called the A- and B- boxes, is able to open up the chromatin (Burnol et al., 1993; Varshney et al., 2015). Once bound, TFIIIC can recruit TFIIIB – composed of the TATA binding protein (TBP), Brf1 and Bdp1 – via an interaction with its Brf1 subunit. Although the presence of a TATA box upstream Pol III-transcribed genes is more of an exception in *S. cerevisiae*, it is widespread in fission yeast and human, providing an alternative and/or complementary mode of recruitment for TFIIIB (Hamada et al., 2001; Pugh and Venters, 2016). Accordingly, *in vitro* studies found that TFIIIC is dispensable for Pol III transcription from TATA box-containing promoters (Dieci et al., 2000).

Once bound, TFIIIB-DNA complexes are stable and are able to recruit Pol III through interactions between Brf1 and Rpc34 (Khoo et al., 2014) and between Bdp1 and Rpc128 (called Rpc2 in fission yeast) (Hu et al., 2015). The interaction between Pol III and TFIIIB causes structural rearrangements within Pol III that activate the TFIIE-like Rpc82/Rpc34/ Rpc31 subcomplex to initiate the opening of double-stranded DNA, ultimately leading to Pol III processive elongation (Abascal-Palacios et al., 2018; Vorländer et al., 2018). During active transcription, TFIIIC is displaced from the gene body (Roberts et al., 2003).

In contrast with the two other RNA polymerases, Pol III does not require auxiliary factors to terminate. Instead, it does so with precision and efficiency on a simple stretch of Ts on the non-template strand (Arimbasseri and Maraia, 2015). The number of consecutive Ts required for efficient termination vary between organisms: four in human, five in fission yeast, and six in budding yeast (Hamada et al., 2000).

After termination, TFIIIB remains bound to the promoter and is able to recruit Pol III again to re-initiate a round of transcription. Pol III re-initiation is especially efficient due (1) to the stable DNA-TFIIIB interaction; and (2) the fact that no other dissociable factors than Pol III and TFIIIIB are required. In fact, it is estimated that 99% of Pol III-dependent transcripts come from re-initiation events (Arimbasseri et al., 2014).

B. Pol III regulation

Pol III transcription is tightly regulated in response to environmental cues, such as starvation (Roberts et al., 2003) or DNA damage (Boisnard et al., 2009). This regulation is central in controlling the amounts of tRNAs and ribosome (due to the transcription of the 5S rRNA by Pol III) within the cells, which determine the cell proliferative capacity. As such, over-expression of Pol III products in mammals has been implicated in tumorigenesis (Marshall and White, 2012). In addition, for unicellular eukaryotes whose environment is exposed to varying conditions, the ability to rapidly stop the synthesis of rRNAs and tRNAs is vital for cell survival and for optimal economy of cell resources (Warner, 1999).

In yeast, the first – and only one known – general repressor of Pol III transcription is the conserved Maf1 protein. Maf1 was first identified in a screen for tRNA-mediated anti-suppression (Boguta et al., 1997). Further studies showed that the lack of Maf1 causes accumulation of pre-tRNAs (Ciesla and Boguta, 2008). Resolved structures of Maf1-Pol III complexes during repression revealed that Maf1 binds the Pol III clamp, causes rearrangement in the Rpc82/Rpc34/Rpc31 subcomplex (important for Pol III initiation), and disturbs Pol III recruitment by TFIIIB (Vannini et al., 2010).

In permissive conditions, Maf1 localizes in the cytoplasm, preventing its repressive activity. In consequence, to repress Pol III, Maf1 has to shuttle from the nucleus to the cytoplasm, a process controlled by Maf1 phosphorylation state. Several central signaling pathways including the TORC1, Pkc1 and CK2 pathways link external and internal cues (such as the absence of DNA damage and nutrient availability) to Maf1 phosphorylation, which prevents Maf1 from entering the nucleus. In contrast, in response to unfavorable conditions, Maf1 is dephosphorylated and is allowed to enter the nucleus, where it can actively repress Pol III transcription – reviewed in (Boguta, 2012).

Maf1-dependent repression is extremely efficient and rapid as it takes place within 15 minutes of starvation (Roberts et al., 2003). After inhibition, Pol III occupancy rapidly decreases. In contrast, TFIIIC occupancy increases, possibly to preserve the genes from encroaching nucleosomes even under periods of transcriptional repression (Arimbasseri et al., 2014).

To crudely summarize our current knowledge of Pol III regulation in yeast, it could be assimilated to a simple, yet very efficient ON/OFF button called Maf1. In mammals however, the situation is more complex, perhaps to accommodate gene-specific regulation of Pol III transcription. Indeed, not all Pol III genes are transcribed in mammalian cells, and the individual Pol III genes transcribed vary depending on the tissue or cell type (Dittmar et al., 2006). For instance, 26% of the tRNA genes that are occupied by Pol III in T-cells are not occupied in HeLa cells (Barski et al., 2010). How this more complex regulation of Pol III transcription is managed within mammalian cells is not well understood yet, but certainly involves additional transcription factors such as the proto-oncogene p53 and the c-Myc factor (White, 2011).

C. Maturation of pre-tRNAs.

After transcription, nascent pre-tRNAs undergo extensive maturation steps before becoming fully functional mature tRNAs. For clarity, the main steps of maturation will be presented hereafter sequentially in their most common order (**figure 0**). However, it has become clear that the pathways of maturations are multiple and non-linear, with many quality control checkpoints, rescue pathways and alternative routes – reviewed in (Hopper and Huang, 2015). In addition, I will not cover here the daunting complexity of the multiple post-transcriptional nucleotide modification deposited on tRNAs at various stages of tRNA biogenesis – recently reviewed in (Han and Phizicky, 2018).

i. 5' leader removal

After transcription, the nascent pre-tRNA is transcribed in a form that contains a 5' leader sequence and a 3' trailer sequence that will not be part of the mature tRNA. The first step of maturation generally consists in the 5' leader removal by the RNase P ribonucleoproteic complex that is essential for cell viability (Xiao et al., 2001).

ii. 3' trailer removal

Three conserved players are involved in the 3' trailer removal: the RNaze Z endonuclease, the 3'-5' exonuclease Rex1 and the tRNA chaperone Sla1 (homolog of human La). Sla1 binds the 3' end of tRNAs and maintains them in a conformation that facilitates endonucleolityc cleavage of the trailer by the RNaze Z protein (Van Horn et al., 1997; Yoo and Wolin, 1997). In competition with Sla1 binding at pre-tRNAs 3' ends, the Rex1 exonuclease constitutes an alternative route for the removal the 3' trailer (Copela et al., 2008). Most tRNAs utilize both pathways for their maturation, although the endonucleotlytic cleavage is the most frequently used, especially with longer (up to 26 nucleotides in *S. cerevisiae*) trailer sequences (Skowronek et al., 2014).

iii. CAA addition

Following the removal of 3' trailer, the maturation of a pre-tRNA requires the addition of a CCA trinucleotide to its 3' end. Intriguingly, for certain ill-shaped unstable pre-tRNA precursors, a second CAA is added at the 3' end (CAACAA), which targets the pre-tRNA for degradation and therefore constitutes a quality control mechanism (Betat and Mörl, 2015).

iv. Splicing

After the maturation of the ends of intron-containing pre-tRNAs, the next step (that does not concern intron-less tRNAs) is the splicing of the introns by the conserved SEN complex. In both budding and fission yeast, this complex is – perhaps surprisingly – localized at the outer mitochondrial membrane (Wan and Hopper, 2018). Therefore, while the other maturation steps occur in the nucleus, pre-tRNA need to be exported in the cytoplasm to be spliced. Providing a checkpoint for pre-tRNA proper maturation before export, the major pre-tRNA exportin Los1 preferentially exports end-matured tRNAs to the cytosol (Chatterjee et al., 2018).

D. Interplay between Pol II and Pol III transcription

i. Pol III transcription affecting Pol II transcription

Work in *S. cerevisiae* revealed that active Pol III transcription can exert negative transcriptional regulation to neighboring Pol II genes (Hull et al., 1994). The mechanism behind this *tRNA gene-mediated silencing* (tgm) is not perfectly clear yet, however, the subnuclear localization of the tRNA gene to the nucleolus is necessary (but not sufficient) for the silencing to occur (Pratt-Hyatt et al., 2013). This suggests that sequestrating Pol II genes within the nucleolus creates an environment unfavorable for efficient Pol II transcription.

In *A. thaliana*, the expression of a pol II-transcribed gene was shown to be negatively correlated with the expression of a Pol III-transcribed gene encoded on the opposite strand of the same DNA fragment (Lukoszek et al., 2013). Similarly, a Pol III-transcribed gene embedded within the first intron of the Pol II-transcribed gene encoding the Rpc5 Pol III subunit represses Pol II transcription through transcriptional interference, providing an elegant feedback loop to control *rpc5* production (Yeganeh et al., 2017).

ii. Pol II transcription affecting Pol III transcription

Recently, Pol II was reported to bind tRNA genes in *S. pombe* (Castel et al., 2014) and in mammals proportionally to Pol III occupancy (Barski et al., 2010; Listerman et al., 2007; Moqtaderi et al., 2010; Oler et al., 2010; Raha et al., 2010). In mammals, Pol II peaks about 200 bp upstream of the Pol III initiation site, where, in most cases, there is no annotated Pol II transcription units. In addition, the histone flanking tRNA genes are modified in ways similar to that of Pol II transcribed genes. For instance, H3K4me3, which is the hallmark of active Pol II transcription (deposited following the Ser5-P-dependent recruitment of Set1), can be found at actively transcribed tRNA genes (Barski et al., 2010).

The strong correlation observed between Pol III and Pol II occupancy and Pol II-related marks suggests a functional relationship between the two polymerases that remains to be elucidated. Pioneer studies suggest that Pol II acts as a Pol III enhancer because Pol II inhibition by α-amanitin reduced the transcription of a subset of Pol III genes (Barski et al., 2010; Listerman et al., 2007; Raha et al., 2010). However these findings are difficult to interpret since Pol III is also sensitive to α-amanitin at high dose, and Pol II inhibition over long time-lapses is likely to generate undirect effects (White, 2011). Currently, the functional relationship between Pol II and Pol III at tRNA genes is still unclear.

E. Aim of the work

Starting with the surprising observation that the loss of Ser2-P strongly affects tRNA expression in fission yeast, we aimed to investigated the mechanism behind this relationship between Pol II transcription and Pol III output. Specifically, we investigated whether Ser2-P and active Pol II transcription could occur directly on tRNA genes,

confirming and expending the puzzling observation that Pol II is present on Pol III-transcribed genes in mammalian cells.

Another important aspect of this work was to assess how the relationship between Pol II and tRNA expression connects to state of the chromatin. In particular, the involvement of Pol III, TFIIIC, TFIIIB and the RSC remodeling complex were tested.

Then the biological relevance of the connection between Ser2-P and tRNA transcription was investigated in regards to Pol III regulation. Good evidences were found that a Maf1-independent repression of Pol III transcription exists, a repression that could be Ser2-P dependent. In consequence, we set up a screen to uncover new regulatory layers of Pol III transcription, highlighting a connection with the SAGA complex that further connects Pol II and Pol III transcription.

2. Results & Discussion

A. A global view on transcriptional changes upon Ser2-P loss.

i. Ultra-deep sequencing reveals a global increase of cryptic transcription in the S2A mutant

In a previous study from our laboratory, the impact of the loss of Ser2-P on the *S.pombe* transcriptome was examined using the microarray technology (Coudreuse et al., 2010). This seminal study revealed the gene-specific requirement of Ser2-P on a genome-wide scale, but due to the technical limitations of the microarrays – mainly, the fact that expression measurement was dependent on human-designed probe, restricting the analysis to the annotated genes, that were, at the time, mostly protein-coding genes –, only a fraction of the transcriptome was analyzed (Malone and Oliver, 2011).

In this study, we used state-of-the-art total RNA-seq technologies to investigate in more details the potential roles of Ser2-P in non-coding RNA (ncRNA) transcription. This technology allows, thanks to small modifications to the Illumina TruSeq stranded RNA-seq protocol (see material & methods), to sequence all stable RNAs longer than 50 nts from the cells. We sequenced two replicates of the wild-type and *S2A* mutant strains to an average depth of coverage of ~300x, providing enough power to detect differences of expression even in lowly expressed genes (**table 1**).

Table1: Conditions studied by RNA-seq.

strain	replicate	depth[a]
wild-type	1	292
wild-type	2	391
S2A	1	285
S2A	2	314

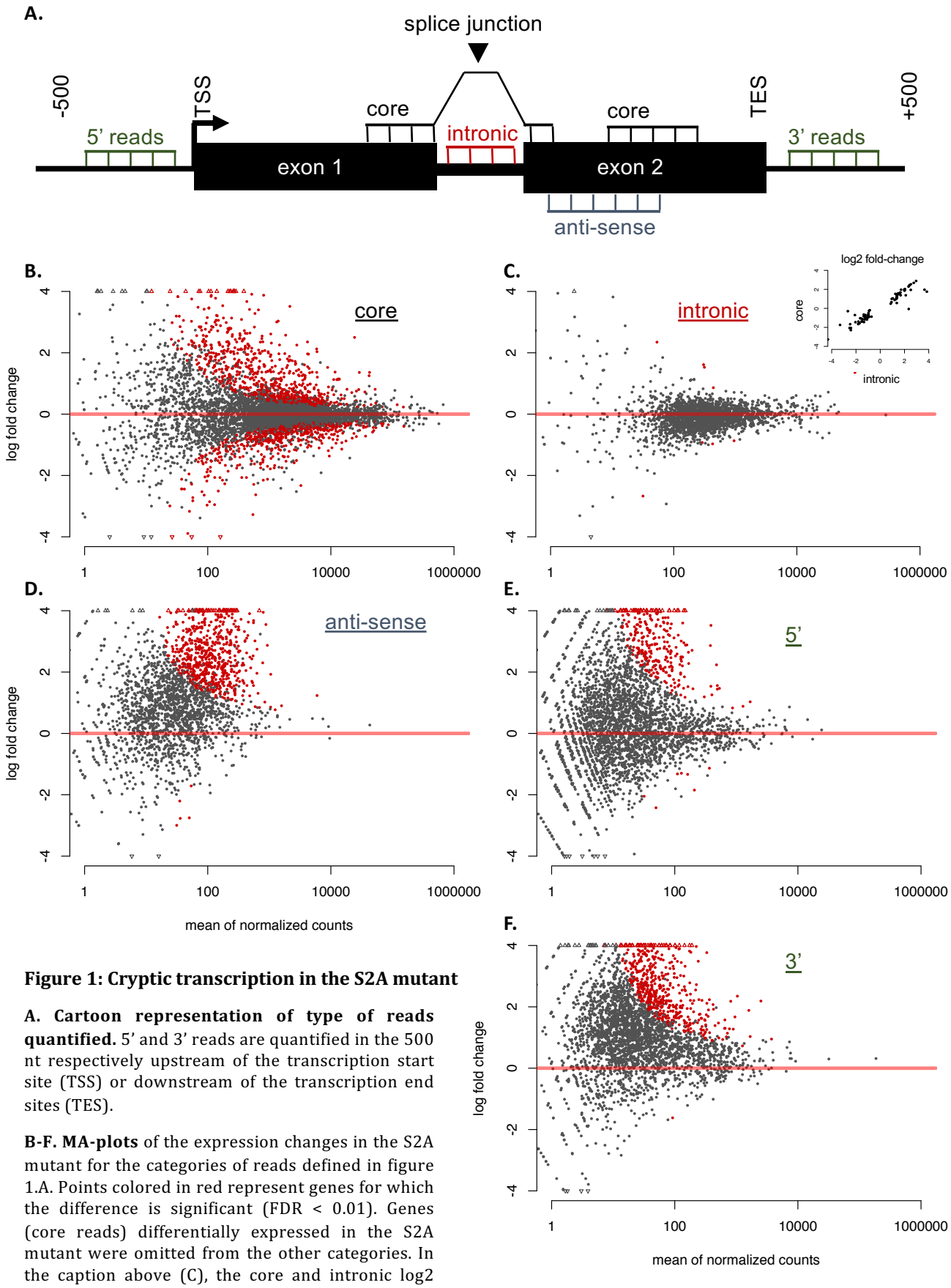


Figure 1: Cryptic transcription in the S2A mutant

A. Cartoon representation of type of reads quantified. 5' and 3' reads are quantified in the 500 nt respectively upstream of the transcription start site (TSS) or downstream of the transcription end sites (TES).

B-F. MA-plots of the expression changes in the S2A mutant for the categories of reads defined in figure 1.A. Points colored in red represent genes for which the difference is significant (FDR < 0.01). Genes (core reads) differentially expressed in the S2A mutant were omitted from the other categories. In the caption above (C), the core and intronic log2 fold change are plotted for the genes whose intronic region is differentially expressed.

a. The depth is calculated based on the number of uniquely mapped reads to the *S. pombe* genome divided by the length of the genome size (14 x 10^6 nt) multiplied by the read length (50 nt).

While the interpretation of RNA-seq data is still dependent on the gene annotation, it is possible to measure expression beyond annotated transcripts. In this analysis, we considered the following regions (**figure 1.A**):

- *core*: this region consists in the exons of the annotated transcripts. Reads mapped to it contribute to what is generally called "gene expression".
- *5':* the 500 nt upstream of the annotated transcripts.
- *3':* the 500 nt downstream of the annotated transcripts.
- *intronic*: this consists in reads spanning introns. About 43% of *S. pombe* protein coding genes have introns (Wood et al., 2002) which are generally much shorter than in higher eukaryotes with an average length of 83 nt (Wood et al., 2012).
- *antisense:* since our RNA-seq is strand-specific, it is possible to distinguish signal arising from the template strand of the annotated genes (core and intronic reads), but also the signal arising from the opposite strand, antisense of the annotated genes.

After normalization based on the median of ratio (Anders and Huber, 2010) for the core regions (applied to all regions), our analysis revealed that 977 genes (14% of the annotated transcriptome) are differentially expressed on their core regions (**figure 1.B**, **table 2**). They will be discussed further below.

Excluding those differentially expressed genes, almost no gene (9) have its introns differentially expressed (**figure 1.C**). Note that when the genes differentially expressed on their core regions where included in the analysis, the number of genes differentially expressed on their intronic regions raised to 82, but the fold change on the introns strongly correlated with the fold change on the exons (see caption above figure 1.C), suggesting that the differential expression on the introns is due to a general change in the gene expression. In contrast, with a *bona-fide* splicing defect, we would expect to observe intron retention without an increase in gene expression.

In addition, only 237 out of the 977 genes differentially expressed had introns, which is actually less than the global proportion of genes with intron (2553/7000). Together, these results argue against an global role for Ser2-P in coordinating splicing in *S. pombe*, which is in agreement with a recent study that found an absence of intron retention in the S2A mutant in fission yeast using splicing-sensitive microarrays (Inada et al., 2016) but contrasts with early studies demonstrating splicing defects of specific reporter introns in human and chicken cells containing S2A mutations (Gu et al., 2013; Hsin et al., 2014). A possible explanation for this discrepancy would be that Ser2-P is important for the co-transcriptional splicing of long introns, which is usually not the case in fission yeast with an average intron length of 83 nt (Wood et al., 2012) as opposed in higher eukaryotes such as human, where the average intron length is of 3356 nt (Lander et al., 2001).

Table2: Regions up- or down- regulated in the S2A mutant

regions	up	down
core	616	361
intron	5	4
5'	536	1
3'	295	7
antisense	652	5

While there were few effects in the intronic regions, many regions in antisense (**figure 1.C**), upstream (**figure 1.D**) and downstream (**figure 1.E**) of the genes were found significantly up-regulated in the S2A mutant (**table 2**). Although the variability is high – probably in part because of the relatively low number of reads mapping to those regions – the distribution of the fold change reveals a global shift of the expression. This suggests a role for Ser2-P in enforcing transcriptional fidelity and preventing both extragenic and anti-sense intergenic "cryptic" transcription, a role that was never demonstrated *per se* until now. However, it could have been anticipated given the well described function of Ser2-P in recruiting the Set2 histone methyl-transferase (Li et al., 2002) (Li et al., 2003) (Xiao et al., 2003) whose activity favors histone deacetylation and represses cryptic Pol II transcription (Carrozza et al., 2005) (Keogh et al., 2005) (McDaniel et al., 2017).

Along with those effects (or lack of) on the extra-exonic regions, the 993 differentially expressed genes (DEG) were affected differently depending on the type of gene considered (**table 3**). For the protein-coding genes, the DEG (522 DEG, about 10% of all protein-coding genes in *S. pombe*) are evenly distributed between up- and down-regulated genes. Consistently with previous publications, *ste11* and *mei2* are included in the down-regulated genes (Coudreuse et al., 2010) and there is an excellent agreement with the 192 Ser2-P-dependent genes previously identified by another group using microarray (Shuman 2016), since 182 of those genes (95%) were also found differentially expressed in our study.

In contrast, long non-coding RNAs (lncRNAs) are mostly up-regulated in the S2A mutant. For the lncRNA genes, 277 were up-regulated from a total of 1540 annotated lncRNA, most of them being antisense RNAs: 227 from a total of 695 annotated antisense RNAs, which constitutes a very significant enrichment (Fisher's exact test p-value < 2.2×10^{-16}). This result is consistent with our initial description of a global up-regulation of the anti-sense non-coding transcriptome in the S2A mutant (**figure 1.C**), probably via the mis-recruitment of the H3K36 methyl-transferase Set2 as discussed above.

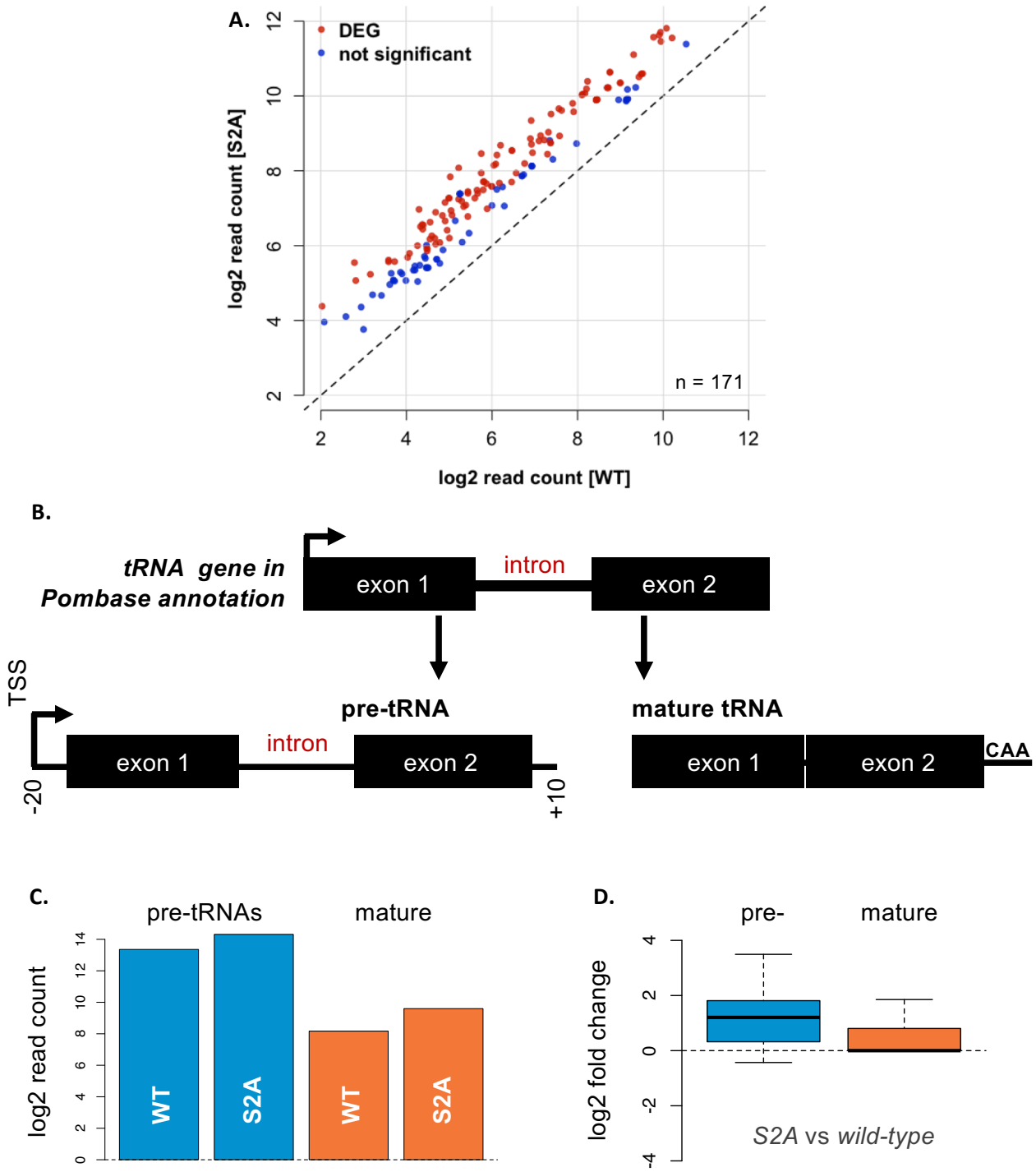


Figure 2: tRNAs over-expression in the S2A mutant.

A. Scatter plot of the log2 read counts for tRNA genes (averaged over two replicates).

B. Cartoon representation of the tRNA gene models used for read quantification. From the Pombase gene models, we extended the region 20 nt upstream and 10 nucleotides downstream in order to generate the 5' and 3' trailer of the precursor tRNAs (pre-tRNAs) gene models. In parallel, we removed the introns and added the CAA trinucleotide to generate the mature tRNA gene models.

C. Log2 total read count on pre- or mature tRNAs (summed over two replicates). The reads were mapped to both gene models in competition and only uniquely mapped reads where counted.

D. Boxplot of the log2 fold-change for pre- or mature tRNAs.

Table3: Type of the 993 genes differentially expressed by the S2A mutation

gene type	up	down
protein-coding	237	285
lncRNA	277	71
snRNA	0	0
snoRNA	3	1
tRNA	91	0

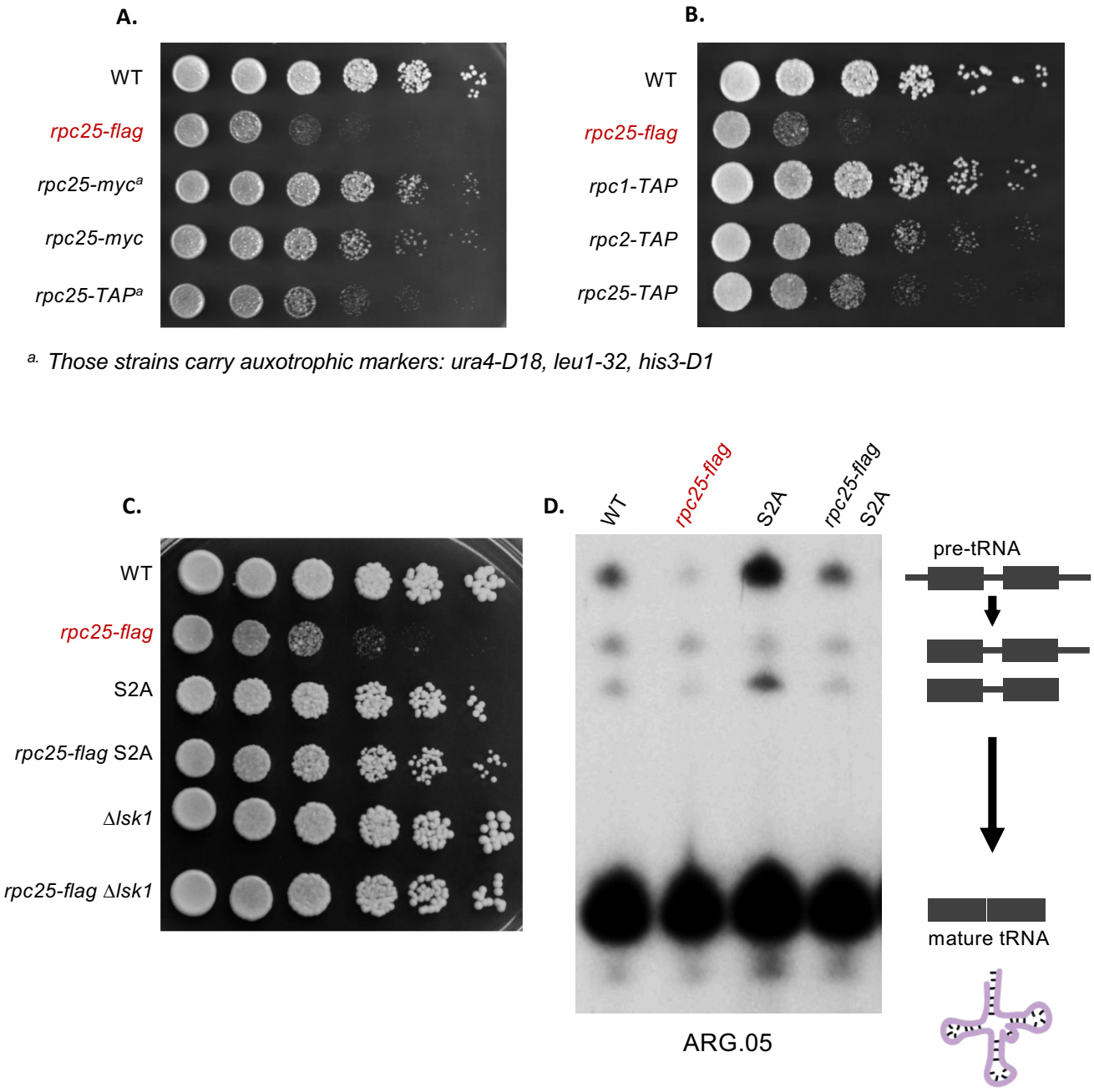
As for the snRNA and snoRNA genes, only 0/7 and 4/55 are respectively found differentially expressed, arguing against a global role of Ser2-P in controlling bulk gene expression for these gene type.

ii. Pre-tRNAs are over-expressed in the S2A mutant.

Finally, to our surprise, we found that all 171 nuclear tRNAs are up-regulated in the S2A mutant, including 91 genes for which the up-regulation is significant (**figure 2.A**). This result is rather unexpected given that those genes are normally transcribed by Pol III and that their expression should not, *a priori,* be affected by mutations within Pol II CTD.

Furthermore, the interpretation of this result is difficult because high-throughput sequencing of tRNAs, especially the steps of primer hybridization or ligation and reverse transcription, is hampered by the very stable and compact secondary structures adopted by the tRNAs (Beltchev et al., 1976) and their high level of modification (13 modifications per molecule on average) (Pan, 2018). Methods have been developed to facilitate tRNA sequencing by removing the modifications (Wilusz, 2015; Zheng et al., 2015) or by partial hydrolysis of the tRNAs (Arimbasseri et al., 2015; Gogakos et al., 2017), but they were not applied to this study. Therefore, it is possible that the reads that map to the tRNA genes do not represent mature tRNAs but rather come from hypo-modified tRNA precursors.

To explore this possibility, we computationally extended the tRNA gene models available in *pombase* (Wood et al., 2012) into either full length tRNA precursors or fully processed mature tRNAs (**figure 2.B**). Then, reads from our RNA-seq experiment were mapped to these gene models in competition, and uniquely mapped reads were counted in order to differentially quantify pre-tRNAs from mature tRNAs. Validating the soundness of our procedure, a conceptually similar method have since been independently published (Hoffmann et al., 2018). Results from our quantification reveal that most of the reads counted on tRNA genes arise from pre-tRNAs (**figure 2.C** – *in log2 scale*), confirming that mature tRNAs cannot be efficiently sequenced without using specific techniques. Both mature and pre-tRNAs are over-represented in the S2A mutant; however, this upregulation is more robust and stronger for pre-tRNAs (**figure 2.D**). Those results were confirmed later using Northern blots that allow to unambiguously discriminate between tRNA precursors, intermediates, and mature forms (see figure 3.D for instance).



a. *Those strains carry auxotrophic markers: ura4-D18, leu1-32, his3-D1*

Figure 3: *Rpc25-flag* is a hypomorphic Pol III allele, suppressed by the loss of Ser2-P.

A-C. Growth assay of the indicated strains. Precultures were grown overnight then diluted in liquid YES media to OD 0.2 and incubated with agitation at 32°C until OD 0.5. From there, 5-fold dilutions were spotted on YES-agar plates and incubated during 3 days at 32°C.

D. tRNA northern blot targeted against SPBTRNAARG.05. On the right, cartoon representations of the various tRNA isoforms recognized by the probe.

B. Genetic interactions connect Ser2-P of the RNA polymerase II with Pol III transcription

To connect the surprising accumulation of (pre-)tRNAs in the S2A mutant with changes in Pol III occupancy on chromatin (an issue more thoroughly discussed later in this thesis), we decided to tag a Pol III subunit in order to assess Pol III distribution genome-wide in ChIP-seq experiments. The use of tagged strains is common in yeast research as creating the tagged fusion gene is generally much easier/cheaper/faster than generating an antibody of satisfying specificity. However, this strategy has the significant drawback that the added tag can, sometimes, affect the protein function. In the case discussed below, the addition of a specific tag on the Pol III subunit Rcp25 leads to a strong growth defect, that is, to our surprise, fully suppressed by the loss of Ser2-P. While such phenotype makes the tagged strain unfit to study Pol III occupancy, we took it as an opportunity to learn more about the connection between Pol II Ser2-P and Pol III transcription.

i. rpc25-flag is an hypomorphic Pol III allele, suppressed by the loss of Ser2-P.

Rpc25 was *a priori* a good candidate for Pol III tagging. This rpb7 paralog constitutes, with its heterodimer partner rpc17 (Ehara et al., 2011), the stalk of the polymerase (Vannini and Cramer, 2012) and its C-terminal end is localized at the periphery of the Pol III complex (Hoffmann et al., 2015), making it potentially more resilient to the conformational constraints brought by the tags. However, tagging of this subunit causes mild to serious growth defect depending of the tag used: *rpc25-myc* grows almost like a wild-type strain, *rpc25-flag* has a marked growth defect, and *rpc25-TAP* an intermediate growth defect (**figure 3.A**). In contrast, TAP-tagging at the C-terminal of the Rpc1 subunit does not lead to a growth defect, which is why we ultimately used this strain for the Pol III ChIP-seq that will be discussed later (**figure 3.B**). This is somehow surprising given that the TAP-tag (21 KDa) is much heavier than the flag (7 KDa for the 5 x flag repeats used) and that Rpc1, the largest Pol III subunit that contains the catalytic site, is at the core of the Pol III complex. Those specific cases illustrate that it is very difficult to predict the functional impact of the addition of a peptide to a protein.

While crossing the *rpc25-flag* strain with the S2A mutant in order to obtain the double mutant *rpc25-flag* S2A and assess Pol III occupancy changes in the S2A, we noticed that the S2A mutation completely suppresses the growth defect of the *rpc25-flag* allele (**figure 3.C**). While the S2A mutant has been studied in the lab (and beyond) for 10 years, this is the only reported occurrence of such a positive genetic interaction, highlighting the importance of Ser2-P in controlling tRNA production. Confirming this result, deletion of the Ser2-P kinase *lsk1* also restores normal growth in the *rpc25-flag* background.

At the RNA level, pre-tRNAs – which we saw are over-expressed in the S2A mutant – are under-expressed in the *rpc25-flag* strain (**figure 3.D**). This is consistent with the idea that the growth defect in that strain is caused by a partial loss of Pol III function. Interestingly, the double mutant *rpc25-flag* S2A restores WT-like levels of pre-tRNAs. This

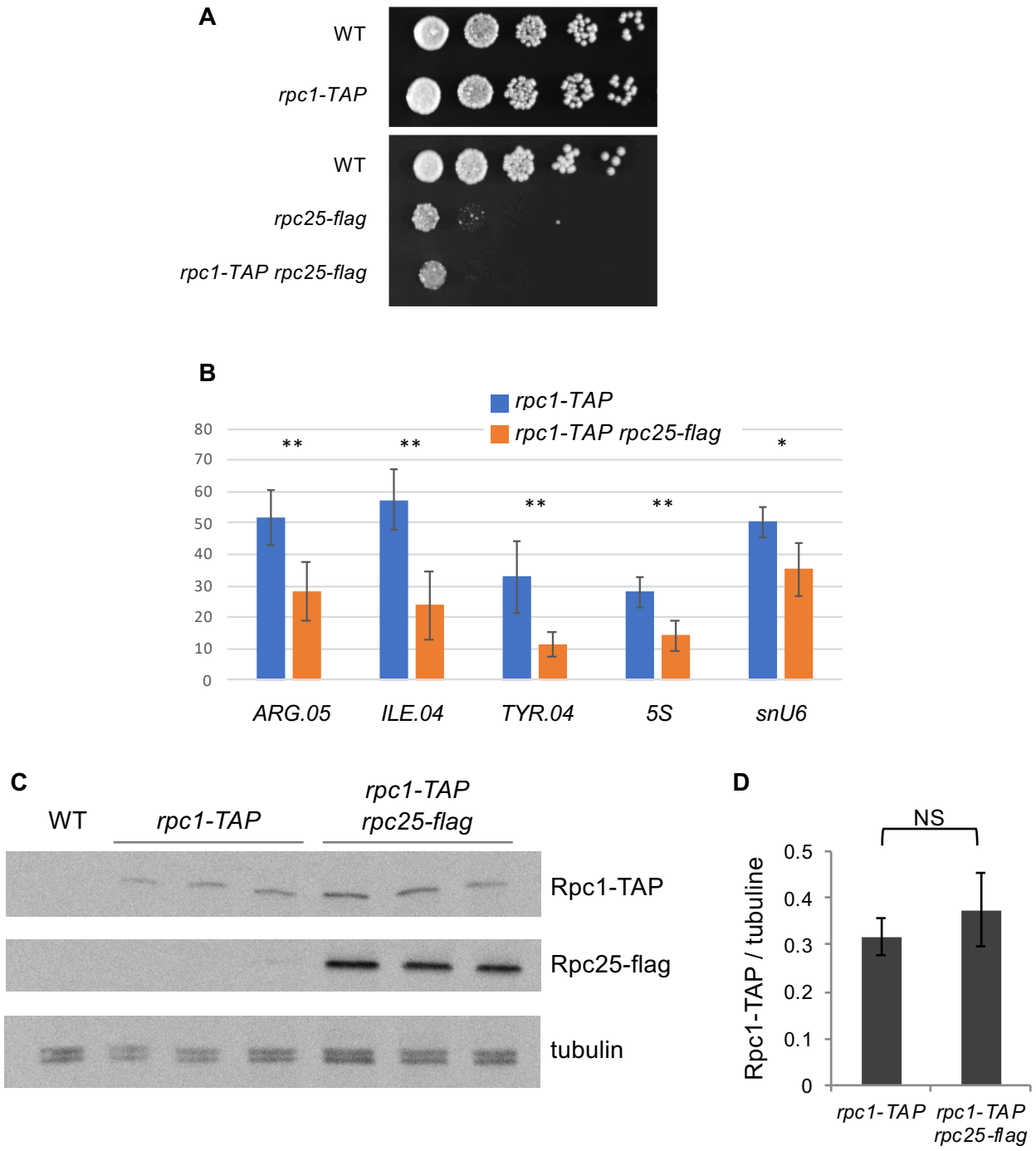


Figure 4: Molecular basis of *rpc25-flag* growth defect

A. Growth assay of the indicated strains. Precultures were grown overnight then diluted in liquid YES media to OD 0.2 and incubated with agitation at 32°C until OD 0.5. From there, 5-fold dilutions were spotted on YES-agar plates and incubated during 3 days at 32°C.

B. ChIP-qPCR experiment on the TAP-tagged Pol III subunit Rpc1. Values are expressed as the percentage of IP over the input, and the error bars represent the 95% confidence interval of the mean over five biological replicates.

C. Western blot of *rpc1-TAP* and *rpc1-TAP rpc25 flag* whole cell protein extracts (using the TCA protocol) in biological triplicates probed against the TAP, flag and tubulin epitopes.

D. Quantification of the chemiluminescent signal in (C.) for Rpc1-TAP normalized against tubulin. The error bars represent the standard deviation of the mean over three replicates. Student's t test is not significant.

correlation between the growth of the strain and the level of pre-tRNAs suggests that the positive genetic interaction between the *rpc25-flag* and S2A alleles is due to compensation of tRNA expression.

As of why the decrease in tRNA species observed in the *rpc25*-flag mutant is more apparent for the pre-tRNA isoforms than for the mature forms, we can hypothesize that the slower vegetative growth of the *rpc25-flag* strain allows the mature tRNAs – which are stables for days (Kadaba et al., 2004) – to accumulate for a longer time in the cells before division. In contrast, the short-lived pre-tRNAs are more directly dependent on transcription as their much smaller half-life (Gudipati et al., 2012) does not allow accumulation.

As a perspective, it would be interesting to assess whether other hypomorphic alleles of Pol III subunits are compensated by the S2A mutation. To our knowledge, no such allele have been described in *S. pombe*, but several mutants in *S. cerevisiae* with point mutations in the highly conserved catalytic subunit of Pol III, Rpc160 (Rpc1 in *S. pombe*), have a slow growth phenotype (Dieci et al., 1995). Among these mutants, *rpc160-112* have been characterized the most and it was elegantly shown that its Pol III transcriptional defect is mainly due to a reduced Pol III elongation rate. However, despite the high conservation of the catalytic core of Rpc1 in eukaryotes, reproducing this mutation is *S. pombe* is not possible as one of the two residues substituted in that mutant (T506I) is not conserved in *S. pombe*. Nevertheless, other mutated residues that impact growth (but are mechanistically less characterized) are conserved in both yeast species, such as the methionine 517 (499 in *S. pombe*), mutated in isoleucine in the *rpc160-206* strain. Creating the *rpc1 M499I* allele along with the *rpc1 M4991 S2A* double mutant in *S. pombe* could allow to assess whether the S2A-dependent suppression of the *rpc25-flag* growth defect can be generalized to other Pol III transcriptional defects.

ii. Insights into rpc25-flag growth defect.

In order to better characterize the *rpc25-flag* mutant, we took advantage of the phenotypically neutral TAP-tagged version of Pol III largest subunit, Rpc1. This additional tag allows to assess the chromatin occupancy of the Pol III complex (or at least, of its Rpc1 subunit) in the *rpc25-flag* mutant. As expected, while the *rpc1-TAP* strain has a wild-type-like growth (**figure 3.B**), the double *rpc1-TAP rpc25-flag* mutant shows the same growth defect as the simple *rcp25-flag* mutant (**figure 4.A**).

By ChIP-qPCR, we observed that the *rpc25-flag* allele causes a significant ~2-fold reduction of Rpc1-TAP occupancy at the five Pol III-transcribed genes tested (three tRNA genes, snU6 and a 5S rRNA gene) (**figure 4.B**), yet the total level of Rpc1 within the cells was unaffected (**figure 4.C**). This suggests that the interaction of the Pol III complex with the DNA is destabilized, in agreement with the decrease in pre-tRNAs transcripts. As a perspective, it would be interesting to see how the S2A mutation affects the Pol III occupancy reduction observed with the *rpc25-flag* strain. If the S2A mutation suppresses that reduction, then it would indicate that the suppression of the Pol III transcriptional defect in the *rpc25-flag* S2A strain occurs on a transcriptional level (for Pol III).

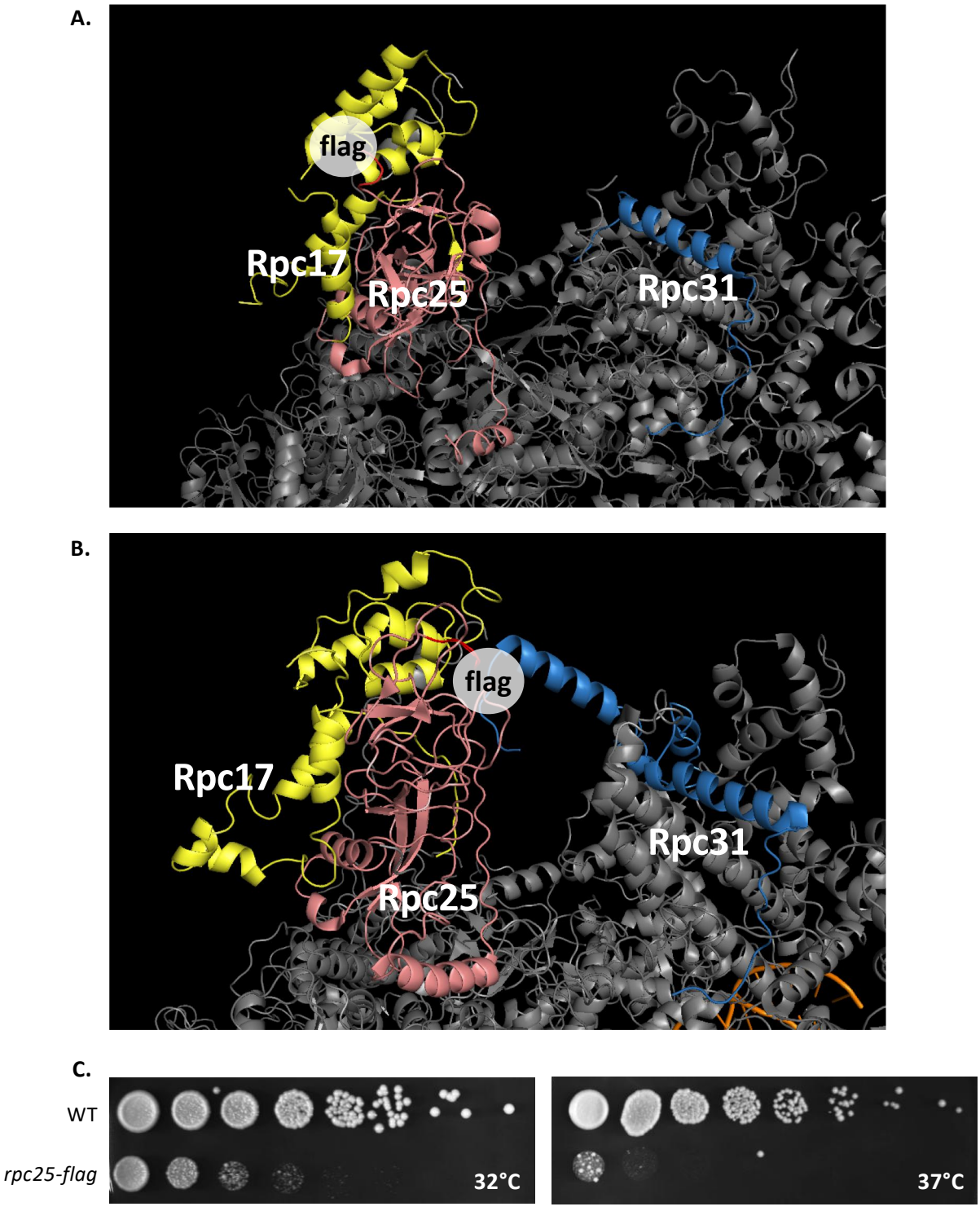


Figure 5: Structural basis of *rpc25-flag* growth defect

A. The Apo pol III open complex structure (protein databank accession number: pdb 6EU2) (Abascal-Palacios, 2018)**.** Rpc17, Rpc25 and Rpc31 are respectively colored in yellow, pink and blue. The 5x flag fusion peptide of the *rpc25-flag* strain is indicated at its expected position at the C-terminal end of Rpc25, which is colored in red.

B. The pol III pre-initiation complex structure (protein databank accession number: pdb 6EU0) (Abascal-Palacios, 2018). The same color code as in (A) was applied, with the addition of orange to color the bound DNA.

C. Growth assay of the indicated strains. 5-fold dilutions were spotted on YES-agar plates and incubated during 3 days at 32°C (left) or 37°C (right).

Work in *Saccharomyces cerevisiae* also highlighted the importance of the well conserved – and essential – Rpc25 subunit. For instance, mutation of a conserved residue within *rpc25* impairs the *in vivo* synthesis of pol III transcripts and causes a temperature-sensitive phenotype (Zaros and Thuriaux, 2005). Through thorough *in vitro* studies of tRNA synthesis with the mutated Pol III complex, the authors showed that the mutation specifically impacts the initiation step of transcription as the tRNA synthesis defect could be relieved by pre-assembling the Pol III initiation complex.

To assess whether the Rpc25-flag fusion protein could impair Pol III initiation in fission yeast, we took advantage of recently published structures of the conformations adopted by Pol III during transcriptional initiation in *S. cerevisiae* (Abascal-Palacios et al., 2018; Vorländer et al., 2018). Because of the conservation of the RNA polymerases among eukaryotes (Huang and Maraia, 2001), those structures should allow a good approximation of the conformations adopted by Pol III in *S. pombe*. In the apo structure (free polymerase), the C-terminal end of Rpc25, where the *flag* peptide was fused, is hanging freely outside of the complex (**figure 5-A)**. However, when the conformational change required for the formation of pre-initiation complex occurs, the C-terminal end of Rpc25 rotates from the outer part of the complex to the inner part, in close contact with Rpc31 (**figure 5-B**). Thus, the addition of a peptide, even as small as the flag, at this location is expected to heavily interfere with the formation of the initiation complex.

Finally, as it is often the case with alleles causing conformational problems, the *rpc25-flag* strain is temperature sensitive: the strong growth defect observed at 32°C is exacerbated at 37°C where almost no growth is observed after three days of incubation (**figure 5.C**).

Taken together, (1) the phenotypes of *rpc25-flag*; (2) the reduction of Rpc1 chromatin occupancy in the mutant; (3) the known function of *rpc25 in* S. cerevisiae*;* and (4) the localization of the expected position of the *flag* peptide at the C-terminal end of Rpc25 in regards to Pol III structures; all point toward the idea that the *rpc25-flag* is a hypomorphic allele that impairs Pol III transcription through destabilization of the pre-initiation complex. However, more mechanistical studies are needed to confirm this idea, as we cannot fully exclude that the phenotypes of the *rpc25-flag* strain are caused by defects at other levels.

In the end, regardless of the causes of the *rpc25-flag* phenotype, its suppression by the S2A mutation provides a useful tool to decorticate the role of Ser2-P in tuning down tRNA expression, a tool that we will use to our advantage in the next chapters.

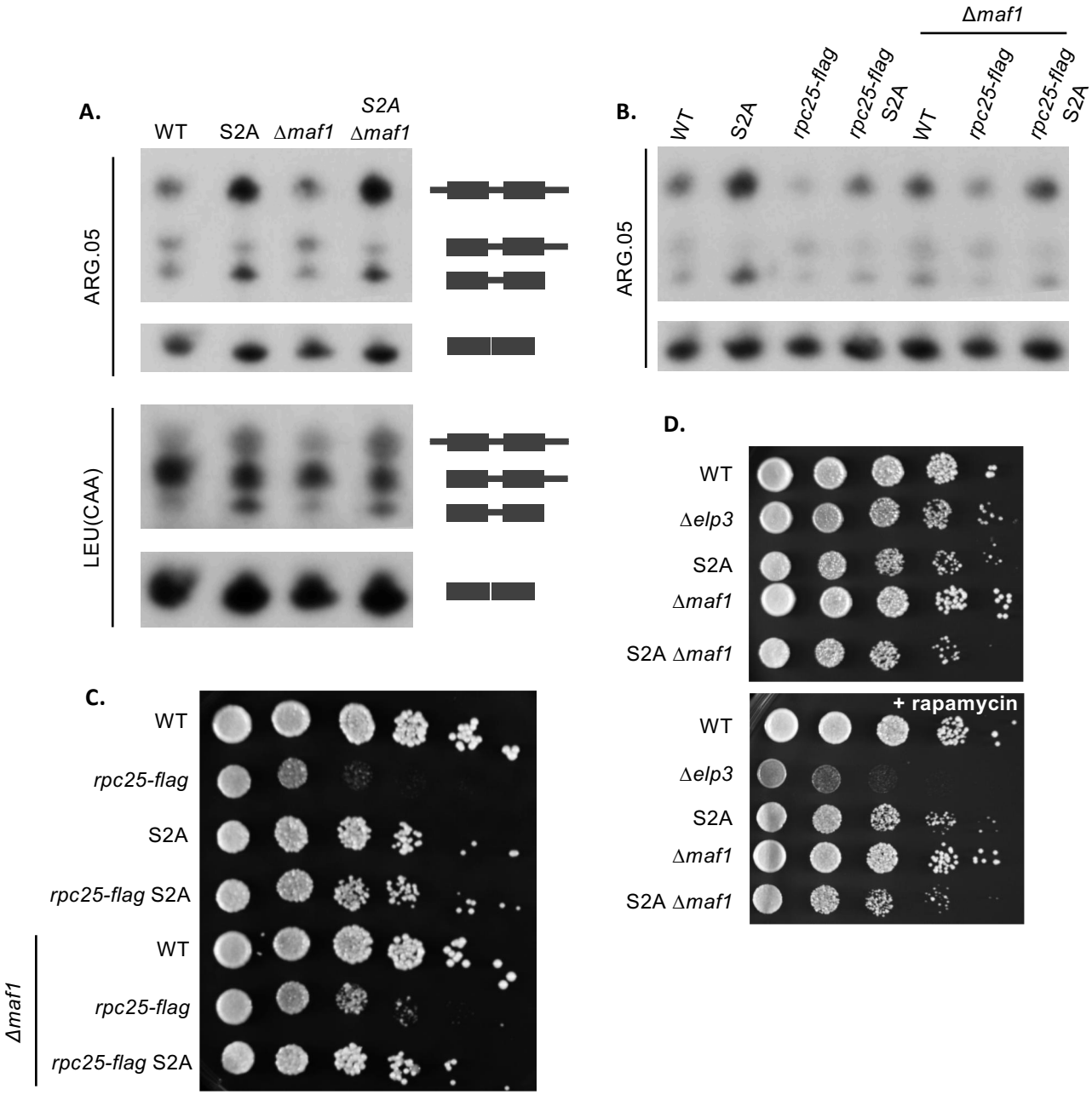


Figure 6: Ser2-P opposes tRNA expression in a *maf1*-independent manner.

A-B. tRNA northern blot targeted against SPBTRNAARG.05 or LEU(CAA) in the indicated strains. On the middle, cartoon representations of the various tRNA isoforms recognized by the probe. For revealing the mature tRNAs (shorter isoform), a much shorter exposure was used.

C-D. Growth assay of the indicated strains. Precultures were grown overnight then diluted in liquid YES media to OD 0.2 and incubated with agitation at 32°C until OD 0.5. From there, 5-fold dilutions were spotted on YES-agar plates and incubated during 3 days at 32°C. In (D), the plates were supplemented with either 20 ng/mL of rapamycin (top panel) or an equal volume of DMSO (bottom panel).

C. Ser2-P opposes tRNA expression in a Maf1-independent manner.

i. Ser2-P opposes tRNA expression in a Maf1-independent manner.

The only known global regulator of tRNA expression in yeast is the conserved Maf1 protein (Ciesla and Boguta, 2008) As such, we wondered whether the increase in (pre-)tRNA transcripts in the S2A mutant was due to a decrease in Maf1 repression. Indeed, even if we did most analysis in rich conditions in which the Maf1 repressor is largely sequestrated in the cytosol, residual Maf1-dependent repression of tRNA expression has been reported in fission yeast in such condition (Arimbasseri et al., 2015).

In our hands, *maf1* deletion barely affects the expression of the tRNAs tested (**figure 6.A**) while the Δ*maf1* S2A double mutant showed a similar increase of pre-tRNA levels as in the simple S2A mutant. Similarly, in the *rpc25-flag* background, *maf1* deletion does not restore pre-tRNA levels, as opposed to the S2A mutation or the double *Δmaf1* S2A mutant (**figure 6.B**). Finally, the growth defect observed in the *rpc25-flag* strain is not fully suppressed by *maf1* deletion, in contrast with the S2A mutant (**figure 6.C**). These results convincingly show that Ser2-P opposes tRNA expression in a Maf1-independent manner.

The fact that *maf1* does not seem to affect pre-tRNA levels in our hands is surprising given the previous reports (Arimbasseri et al., 2015). Still, there is a partial suppression of the *rpc25-flag* growth phenotype (**figure 6.C**), suggesting that either *maf1* deletion affects the expression of some other tRNAs than those we tested, or that our crude assays are not able to detect more subtle changes in pre-tRNA expression.

ii. Ser2-P, a new regulatory layer in tRNA expression?

Given the accumulation of pre-tRNAs observed in the S2A mutant, we wondered whether Ser2-P would be required to repress tRNA expressions in conditions where nutrient availability is limited. This regulation could prove especially relevant since Lsk1, the Ser2-P kinase, is activated by phosphorylation in stress conditions by the Sty1 MAP kinase (Sukegawa et al., 2010).

In *S. cerevisiae*, a shift from YPD medium to YPD 0.15x causes a dramatic, yet reversible, decrease in pre-tRNA levels (Roberts et al., 2003). In fission yeast, a similar shift from YES (the equivalent of YPD medium) to YES 0.15x causes full length precursors levels to drop rapidly, while intermediate forms take up to 75 minutes to fade (**figure 7.A, upper panel**). Surprisingly, despite the abundant literature describing the role of *maf1* in inhibiting Pol III transcription in response to starvation in *S. cerevisiae* and mammals (Ciesla and Boguta, 2008; Desai et al., 2005; Karkusiewicz et al., 2011; Orioli et al., 2016; Roberts et al., 2006), the decrease in pre-tRNAs still occur in *maf1*-deleted cells in our conditions, although perhaps at a slower rate (**figure 7.A, lower panel**). Concerning the lack of impact of *maf1* deletion, it could be more appropriate to test other types of starvations, such as nitrogen starvation, as described {Davie, 2015 #374}.

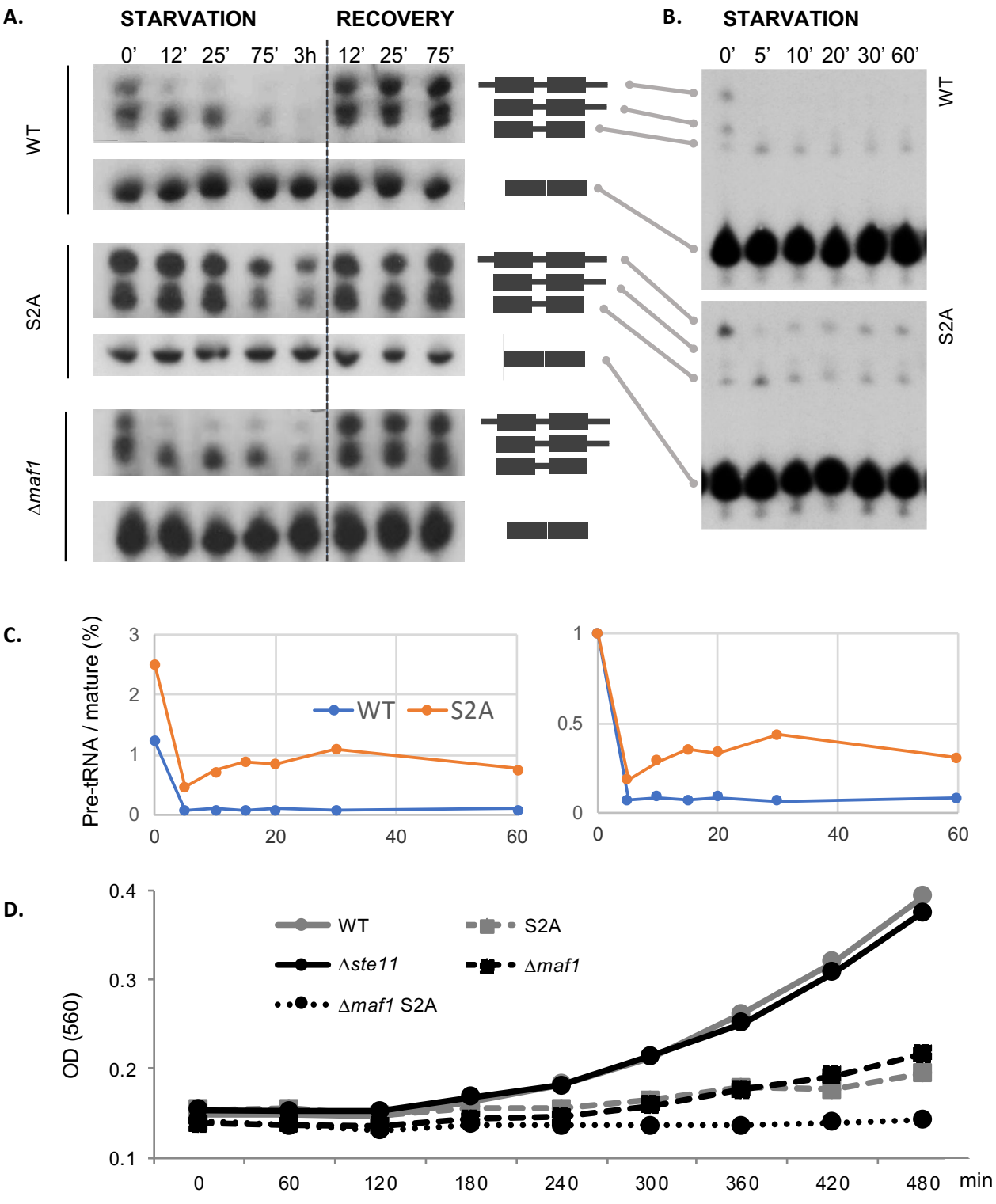


Figure 7: Ser2-P, a new regulatory layer for tRNA expression ?

A-B. tRNA northern blot targeted against SPBTRNAARG.05 in the indicated strains. Cells were growth until OD 0.5 in YES medium then shifted in YES 0.15x (STARVATION) for the indicated times. For (A), the cells were shifted back into YES medium (RECOVERY).

C. Quantification of the full length pre-tRNA for (B). After background subtraction, pre-tRNA levels were normalized to the mature tRNA level. On the right panel, the ratios are scaled to the values at the first time point.

D. Growth assay (OD over time) for the indicated strains diluted at OD 0.15 after overnight free growth leading to quiescence (OD ~9) in YES medium.

This result strongly suggests that, at least in fission yeast and in the tested conditions, another pathway regulates tDNA repression in place of, or redundantly with, Maf1.

To test whether Ser2-P would be involved in such pathway, we did the same shift in the S2A mutant (**figure 7.A, middle panel**), revealing an important delay in pre-tRNA decrease. However, as this preliminary result is not perfectly reproducible, we should be cautious in not over-interpreting it. Nevertheless, repetition of this experiment with shorter timing (and the WT control and S2A mutant processed together with the same batch of media) show that, although the inhibition takes place more rapidly than our preliminary result suggested, the full length pre-tRNA fails to fully fade in the S2A mutant (**figure 7.B**). Indeed, while in the WT the signal for the full length pre-tRNA drops to less than 10% of the signal before the shift, almost 50% of that signal is retained after 30' of starvation in the S2A mutant (**figure 7.C**). This result suggests that Ser2-P is required for efficient tDNA repression upon starvation, a role that might be partially redundant with that of Maf1.

Arguing that both Maf1 and Ser2-P might act in parallel to control tRNA expression, an attentive reader might have notice that, although both the S2A and *Δmaf1* simple mutants show normal growth in spot assays (**figure 6.D**), a small growth defect is observed for the double S2A *Δmaf1* mutant. To assess how this negative genetic interaction affects growth in recovery following long-term starvation, we let fission yeast cells grow into stationary phase overnight. After redilution in YES medium, both the S2A and Δ*maf1* culture showed an extended lag phase before resuming growth, while the double *S2A Δmaf1* mutant did not grow at all during the 8 hours of the experiment (**figure 7.D**). This result suggests that both Maf1 and Ser2-P are important to survive long term starvation and/or reinitiate exponential grow after starvation. This makes sense given that a failure to efficiently shut down transcription and translation in conditions of nutritious stress would jeopardize the chance that a cell remains in good shape to rapidly regrow after the stress, as the cell resources would rapidly be exhausted (Huang et al., 2015).

Altogether, our somewhat preliminary results converge toward the idea that Ser2-P acts in parallel than Maf1 to control pre-tRNA level, although more work will be required to fully validate this hypothesis. For instance, it would be interesting to see how pre-tRNAs levels evolve upon starvation in a Δ*maf1* S2A double mutant.

iii. Δmaf1 is not rapamycin-sensitive

As a side note, we took advantage of our full length *maf1* deletion mutant to clarify a discrepancy in the literature concerning a *maf1*-related rapamycin sensitivity. Indeed, researchers from Gustafsson laboratory reported that *maf1* deletion causes rapamycin sensitivity (Carlsten et al., 2016) while researchers from Maraia's laboratory found that *maf1* overexpression, but not deletion, causes rapamycin sensitivity (Arimbasseri et al., 2015).

This disagreement could be caused by the different genetic background of their *maf1* deleted strains or the use of different rapamycin concentration. In addition, the gene

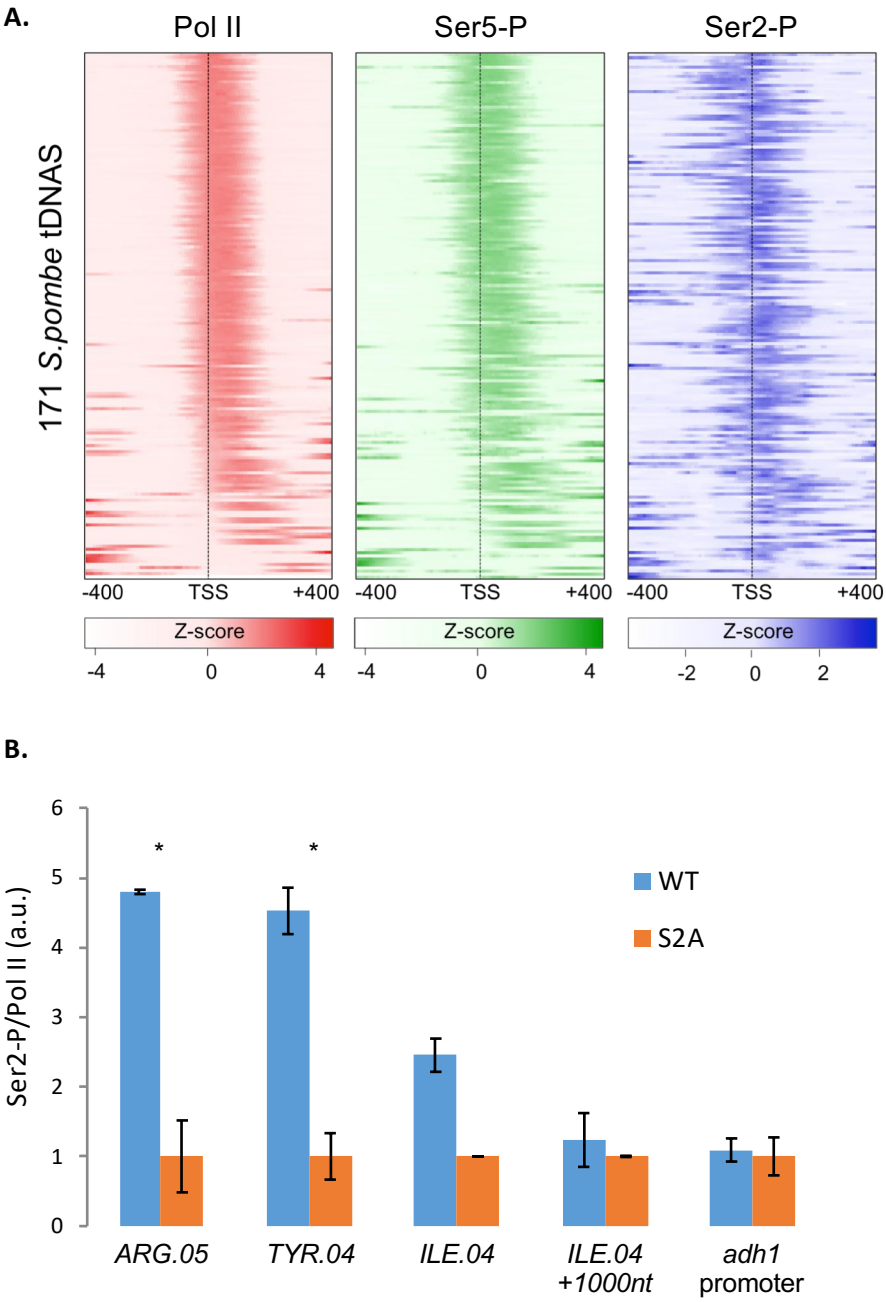


Figure 8: Phosphorylated Pol II associates with tRNA genes.

A. Heatmaps of the Z-score of the ChIP signal for Pol II, Ser2-P and Ser5-P centered on tDNA TSS. The raw data was obtained from Murakami's lab, now published in (Kajitani, 2017).

B. Barplot of the Ser2-P (*ab5095* antibody) ChIP-qPCR signal normalized on Pol II (*8WG16* antibody), scaled to 1 in the S2A condition (background level). Error bars represent the standard error of the mean (SEM) over two biological replicates. Statistically significant differences of mean are indicated by a (*) and were computed with a 2-tailed unpaired t-test assuming unequal variance. ARG.05, TYR.04 and ILE.04 are three tRNAs. ILE.04 +1000 nt refers to an intergenic region located 1000 nt upstream of ILE.04.

annotation for *maf1* was updated in 2009 with the definition of a new ATG start codon 228 nt upstream of the previously annotated ATG (Desai et al., 2005; Wood et al., 2012). Therefore, *maf1* deletion based on the old annotation – such as the *Δmaf1* strain of the Bioneer deletion library (data not shown) – would retain the first two *maf1* exons whose protein product could possibly affect the strain phenotypes in ways that a clean full length deletant does not.

In our hands, the full length deletion of *maf1* in a clean genetic background was not sensitive on plates containing 20 ng/mL of rapamycin, in contrast with our positive control, *Δelp4 (Bauer et al., 2012)* (**figure 6.D**). This absence of phenotype corroborates the findings of Maraia's lab, and makes also more sense in regards to the action of rapamycin within the cells. Indeed, rapamycin inhibits the TOR (Target Of Rapamycin) complex 1 whose kinase activity favours cell proliferation (Cafferkey et al., 1994). In fission yeast, the inhibition is partial, explaining why WT cells are naturally rapamycin resistant in contrast with *S. cerevisiae* (Takahara and Maeda, 2012). In any case, inhibition of the TOR complex 1 causes Maf1 dephosphorylation and nuclear localization, leading to the repression of Pol III transcription. Thus, Maf1 deletion is expected to lessen the impact of rapamycin inhibition while Maf1 overexpression would aggravate it, which is consistent with the absence of rapamycin sensitivity observed in our hands and in Maraia's laboratory.

D. Evidences for Pol II activity at tRNA genes.

While we have established that Ser2-P opposes tRNA expression in a Maf1-independent way and contributes to tRNA down-regulation upon starvation, the mechanisms at play are still elusive. In particular, we wondered whether this putative new role could arise from the direct involvement of Pol II in controlling Pol III transcription.

i. Phosphorylated Pol II associates with tRNA genes.

No gene coding for proteins involved in tRNA transcription was differentially expressed in the S2A mutant. As such, there is no evidence that the observed over-expression of tRNAs comes from undirect effects. Therefore, in order to assess whether a relation of direct causality could explain the observed molecular phenotype, we wondered whether Pol II – and in particular, Ser2-P – associates with tRNA genes.

Through the re-analysis of ChIP-seq data from a previous collaboration with Murakami's lab (Kajitani et al., 2017), we found that Pol II strikingly associates with virtually all nuclear tRNA genes in *S. pombe* (**figure 8.A**). This association is generally centered on the annotated tRNA genes and does not coincide with neighboring Pol II TSS. Importantly, Ser2-P and Ser5-P were also detected, which is consistent with the idea that Ser2-P within Pol II CTD might have a local role on Pol III-dependent tRNA transcription, and that the perturbation of this putative role in the S2A mutant causes tRNA over-production.

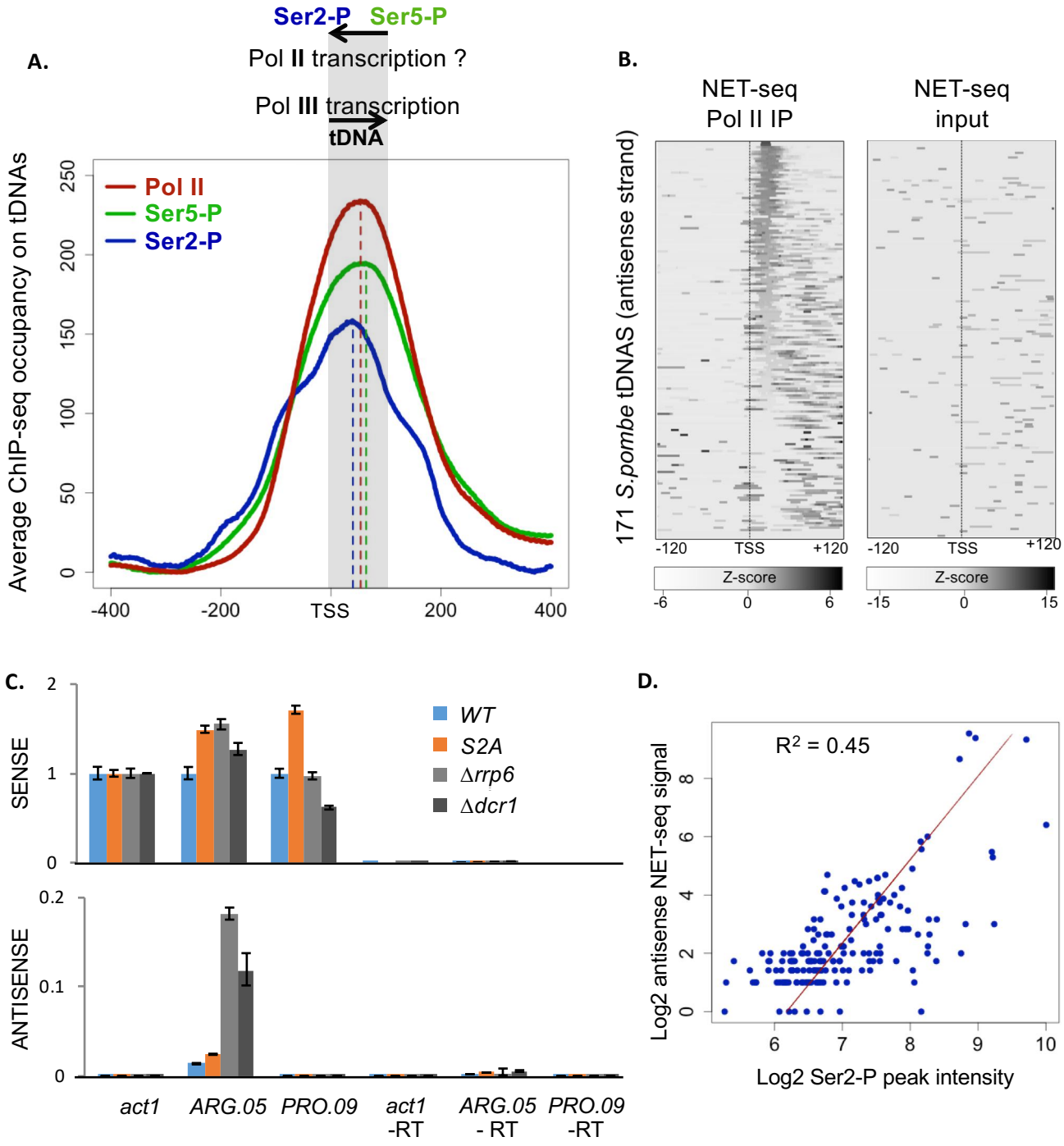


Figure 9: Anti-sense Pol II transcription at tRNA genes.

A. Meta-gene average profile of ChIP-seq signal for Pol II, Ser2-P and Ser5-P centered on tDNAs TSS. The raw data was obtained from Murakami's lab, now published in (Kajitani, 2018).

B. Heatmaps of the Z-score of the NET-seq signal on the antisense strand for Pol II IP or input centered on tDNA TSS. The raw data was obtained from Morillon lab, now published in (Wéry, 2018).

C. Strand-specific RTq-PCR in the indicated strains. Values were normalized to the expression of *act1 SENSE*. ARG.05 and PRO.09 are tRNAs. *–RT* indicates controls for which the reverse transcriptase was not added to the reaction. Error bars represent standard deviation from two technical replicates.

D. Log-log relationship between the level of Pol II-specific antisense transcription (NET-seq) and of Ser2-P at tDNAs. Linear regression is very significant (p-value < 2.2×10^{-16}) with a R^2 coefficient of 0.45.

Using ChIP-qPCR, we were able to confirm the association of Ser2-P normalized on Pol II on 3 tRNA genes, with Ser2-P levels 3- to 5-folds above background (level of Ser2-P in the S2A mutant). In contrast, the level of Ser2-P does not rise above background in an intergenic region or even at the promoter of the housekeeping gene *adh1*.

Altogether, those results show that Pol II associates with tRNA genes in *Schizosaccharomyces pombe*, echoing previous findings in human cell lines (Barski et al., 2010; Moqtaderi et al., 2010; Oler et al., 2010; Raha et al., 2010), mice (Canella et al., 2012; Carriere et al., 2012) and fission yeast (Castel et al., 2014).

ii. Anti-sense Pol II transcription at tRNA genes.

Examining more closely the patterns of Ser5-P and Ser2-P on tRNA genes, we noticed that, on average, the Ser2-P signal extended more upstream of the genes than the – almost confounded – signals of Ser5-P and Pol II (**figure 8.A**). This trend, more apparent on meta-gene profiles (**figure 9.A**), suggests the occurrence of a transcription-dependent transition from Ser5-P (Pol II initiation mark) toward Ser2-P (Pol II elongation mark) in antisense of the orientation of the tRNA genes.

While RNA-seq is a great technique to interrogate the steady-state abundance of RNA, i. e., the result of transcription rate and degradation rate, the recently developed NET-seq method provides an unmatched snapshot of the transcriptional activity of RNA polymerases (Churchman and Weissman, 2011). It relies on the sequencing of the transcribed RNA exiting the polymerase following its co-immunoprecipitation. As the immunoprecipitation can be targeted against either Pol I, Pol II or Pol III, the results are polymerase-specific, which in our case allow to distinguish signal specifically from Pol II transcription.

Confirming our interpretation of the phosphorylated Pol II ChIP-seq data, analysis of Pol II NET-seq data obtained in collaboration with Morillon's lab revealed small nascent Pol II-associated transcripts mapping in anti-sense on tDNAs (**figure 9.B**). They are found in low abundance in comparison with other Pol II transcripts and are probably very short-lived, as they are not detected in the input fraction. Moreover, strand-specific RT-qPCR confirmed the existence of the antisense transcript on one tRNA gene out of the two that were tested (**figure 9.C**). Remarkably, the antisense transcript appeared to be stabilized in the *rrp6* and *dcr1* ribonucleases deletion mutants, suggesting that these two nucleases are involved in their degradation. These results are consistent with previous observation of Dcr1-dependent siRNA mapping in antisense of tDNAs (Castel et al., 2014).

Finally, we found a small (R^2=0.45), yet significant correlation between the level of Ser2-P (elongating Pol II) and the level of antisense transcription detected by NET-seq (**figure 9.D**). This correlation reinforces the idea that the antisense transcription on tDNA is Pol II dependent. Reversely, it argues against the proposed idea that Pol II occupancy on tDNA genes is solely due to the mis-recruitment of Pol II by the TATA box upstream of tRNA genes (Carriere et al., 2012), since Pol II activity is detected in antisense. However, we should note that sense tRNA contamination in both the input and IPed fractions, inherent to the NET-seq protocol (Churchman and Weissman, 2011), impairs the detection of Pol II-specific transcription on the sense direction. Therefore, we cannot exclude that some Pol II transcription on tDNAs also occurs in the sense direction.

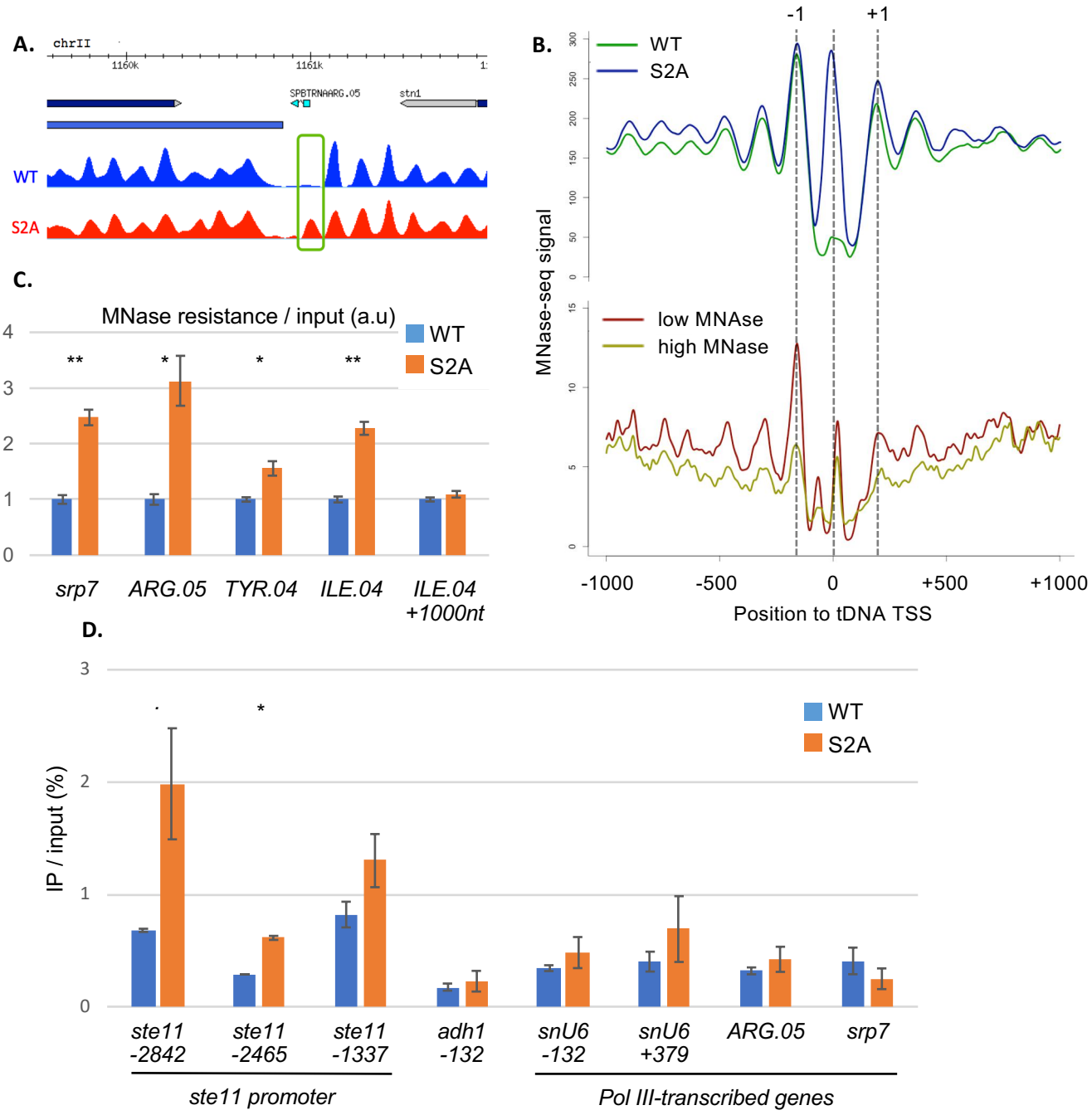


Figure 10: Chromatin changes at Pol III genes.

A. Snapshot of the the MNase-seq profiles in the indicated conditions. Tracks are available on an online genome browser at the following address: https://tinyurl.com/ybdw3bra (Materne, 2015).

B. Meta-gene average profile of MNase-seq signal in the indicated conditions. The data for the WT and S2A conditions are from (Materne, 2015) (upper panel); the data for the high and low MNase conditions are from (Gal, 2016) (lower panel).

C. Nucleosome scanning/MNase-qPCR in the indicated conditions. ARG.05, TYR.04 and ILE.04 are three tRNAs. ILE.04 +1000 nt refers to an intergenic region located 1000 nt upstream of ILE.04. srp7 is another Pol III-transcribed gene.

D. Percentage of IP over input for H3 ChIP-qPCR signal. Error bars represent the standard error of the mean (SEM) over two biological replicates. Statistically significant differences of mean are indicated by a (*) and were computed with a 2-tailed unpaired t-test assuming unequal variance.

E. Ser2-P affects the chromatin structure at tRNA genes.

i. Increased micrococcal nuclease resistance at tRNA genes in the S2A mutant.

Having established that Pol II is transcribing tRNA genes in anti-sense and that the loss of Ser2-P leads to an increase in sense tRNA expression, we wondered whether this differential expression was due to changes at the chromatin level. Indeed, previous work in our lab linked Ser2-P to the RSC chromatin remodeler activity via a modulation of the level of histone acetylation at genes promoters (Materne et al., 2015; Materne et al., 2016). The RSC complex (remodel the structure of chromatin) is an ATP-dependent chromatin remodeler conserved within eukaryotes whose role in the establishment of nucleosome depleted regions at gene's promoters was described by us and others (see **Appendix 1**: *Yague-Sanz et al., 2017 – "A conserved role of the RSC chromatin remodeler in the establishment of nucleosome-depleted regions"*).

In order to assess changes in chromatin structure in the S2A mutant, we reanalyzed the MNase-seq data published along the aforementioned publications, focusing on tRNA genes. MNase-seq is a method that assesses the resistance of the chromatin to digestion by the micrococcal exonuclease. Well positioned nucleosomes typically provide digestion resistance, in contrast with internucleosomal regions and nucleosome depleted regions that are more labile to digestion. Strikingly, tDNAs coincide with regions completely digested in a WT strain while a well-defined peak of MNase resistance is systematically detected in the S2A mutant (**figure 10.A, 10.B – upper panel**). This result was confirmed with independent experiments using nucleosome mapping (a technique that could have been called "MNase-qPCR" by analogy to the ChIP-qPCR/ChIP-seq) on four Pol III genes and one intergenic region as negative control (**figure 10.C**).

ii. Nature of the micrococcal nuclease resistance at tDNA genes.

However, despite the fact that variation in MNase resistance is often associated with differential nucleosome binding, ChIP-qPCR experiments on histone H3 do not provide consistent evidence that the increased in MNase resistance on Pol III-transcribed genes is due to increased nucleosome (or at least, histone H3) occupancy (**figure 10.D**). For this experiment, we used as positive control the *ste11* promoter region that was shown in a previous study to be more occupied by histone H3 in the S2A mutant (Materne et al., 2015), and as negative control, the housekeeping *adh1* promoter. No significant difference was observed for the four Pol III-transcribed loci (including one tDNA) assessed.

As the discrepancy between the ChIP an MNase experiments raises the question of the nature of the MNase resistance, we explored several hypotheses in an attempt to explain it.

(1) Concerns have been raised in the community about the higher sensitivity of some nucleosomes to variation of technical parameters of the MNase-seq experiments.

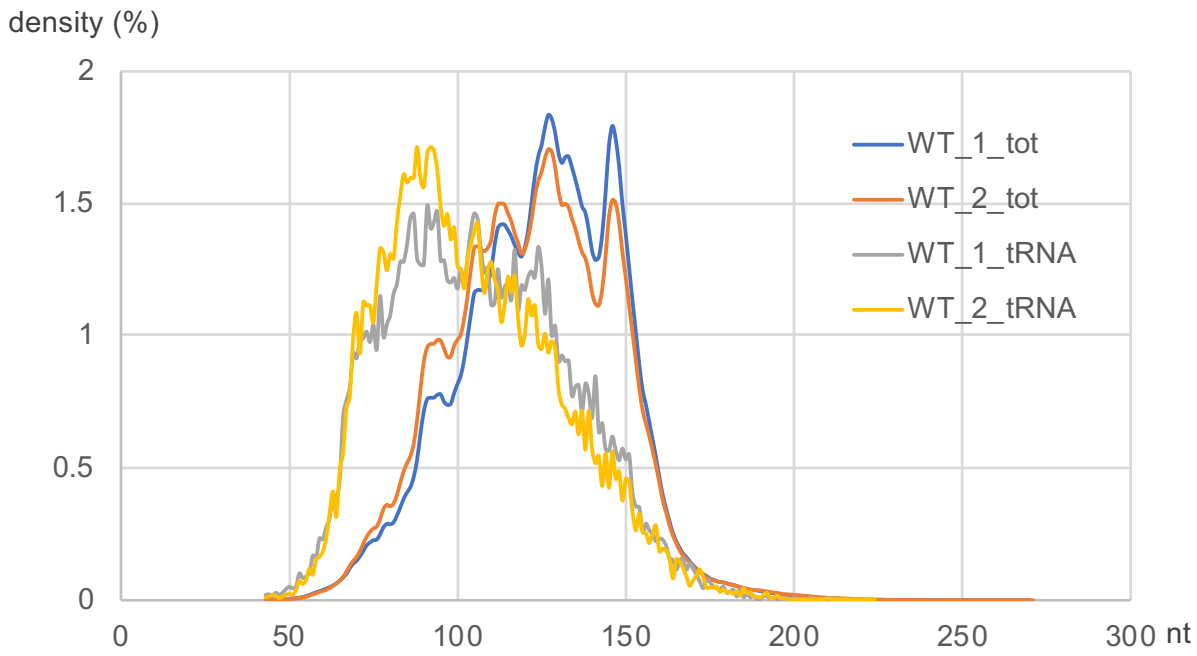


Figure 10-bis: Size of the MNase-resistant particles at tDNAs.

Distribution of the size of the MNase-resistant particles (distance between the left-most end of the left read and the right-most end of its mate) for the WT conditions (two replicates) of the paired-end MNase-seq data described in figure 14.A. Either all genome positions were considered (tot), or specifically the regions located between -100 nt from tDNA TSS to tDNA TES (tRNA).

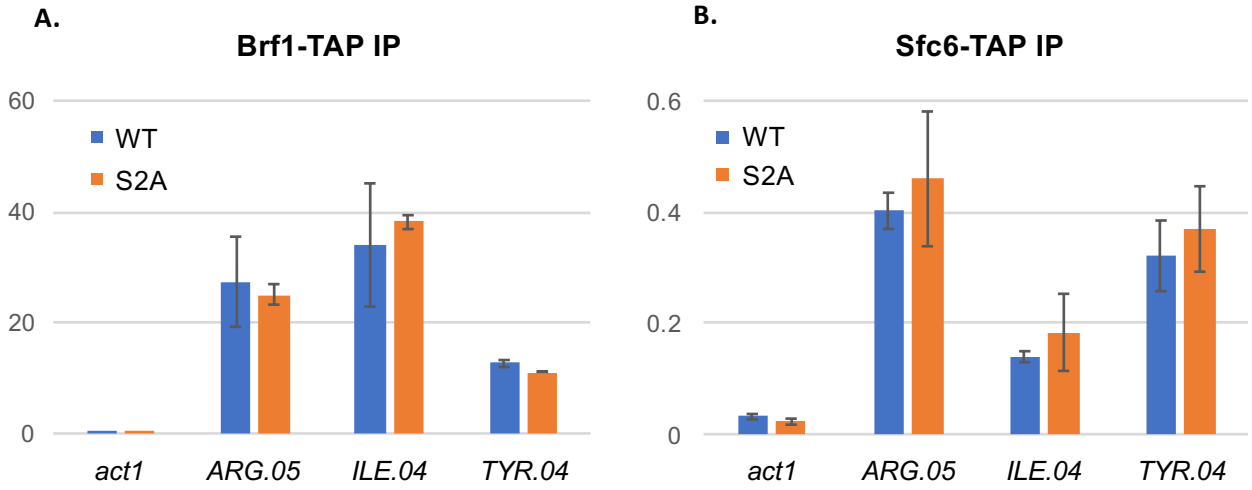


Figure 11: Normal association of Pol III transcription factors to tDNAs in the S2A mutant.

A. ChIP-qPCR experiment on the TAP-tagged TFIIIB subunit Brf1. Values are expressed as the percentage of IP over the input, and the error bars represent the standard deviation of the mean over two biological replicates.

B. ChIP-qPCR experiment on the TAP-tagged TFIIIC subunit Sfc6. Values are expressed as the percentage of IP signal over the input, and the error bars represent the standard deviation of the mean over three biological replicates.

Indeed, depending on the degree of digestion by the MNase, some well-defined peaks of MNase resistance can completely disappear while most other peaks are barely affected (Henikoff et al., 2011; Kubik et al., 2015; Pradhan et al., 2015; Vera et al., 2014). Those "fragile" sites are mainly located at gene's promoters, which are accordingly classified into two classes: those with stable -1 nucleosome (= the first nucleosome before the TSS) and those with fragile -1 nucleosome, which tend to have larger nucleosome depleted regions (Kubik et al., 2015). To assess how the striking difference in MNase protection at tDNAs observed in the S2A mutant is related to this "fragility", we compared the MNase-seq profiles in WT and S2A with WT samples either highly or lowly digested (Gal et al., 2016). The whole regions centered on tDNAs is sensitive the level of digestion, especially the -1 nucleosome (**figure 10.B, lower panel**). In contrast, in the S2A mutant, the -1 nucleosome is largely unaffected, while there is tremendous difference at tDNAs TSS (**figure 10.B, higher panel**). This comparison suggests that the observed difference of MNase resistance between the WT and S2A mutants is not due to technical variation on the degree of MNase digestion and reveals a specific impact of the loss of Ser2-P on the chromatin structure at tRNA genes.

(2) One generally accepted view is that transcription factors (TFs) compete with nucleosomes for DNA binding. In consequence, regions bound by TFs are often nucleosome free and MNase-sensitive (Schones et al., 2008; Zhu et al., 2018). However, this view has been recently challenged by an analysis comparing MNase-seq and histone ChIP-seq data (Chereji et al., 2017). The authors found that the MNase-sensitive particles at yeast promoters (the so-called "fragile nucleosome") do not contain histones and might be composed of transcription factors or chromatin remodeler complexes instead, which are providing partial MNase protection. In particular, they found that the Pol III TFs, TFIIIB and TFIIIC, interact strongly with DNA and provide strong MNase protection. This hypothesis, supported by other studies (Nagarajavel et al., 2013; Shukla and Bhargava, 2018), makes sense in this specific case since it is difficult to reconciliate strong nucleosome occupancy at tDNAs in the S2A mutant with the increased pre-tRNA levels, since the extra nucleosome would be expected to sterically impair Pol III transcription. In addition, the size of the DNA protected by the MNase-resistant particle at tRNA genes is much smaller (with a mode of about 90 nt) to that of MNase-resistant particles throughout the whole genome (*i.e.,* mostly nucleosomes, with a mode of about 120-147 nt), suggesting that their nature differ (**figure 10-bis)**. To test whether the change in MNase sensitivity in the S2A mutant could be attributed to non-histone protein bindings, we compared the occupancy of TFIIIB, TFIIIC and Pol III in the WT and S2A strain. Pol III ChIP-seq data, although complex to interpret (see next point – *iii*), revealed no increase (there is actually a clear decrease) in Pol III occupancy on tDNAs in the S2A mutant (**figure 12**, discussed later). Similarly, TFIIIB (**figure 11.A**) and TFIIIC ChIP-qPCR (**figure 11.B**) respectively targeting their Brf1 and Sfc6 tagged subunits did not highlight any significant change on the three tDNAs tested. All in all, we have no evidence that the peak of MNase resistance we observe in the S2A mutant is due to a change in occupancy in Pol III or its transcription factors. However, we cannot exclude that the affinity of these factors with the DNA is increased in the S2A mutant, affecting the MNase resistance but not their occupancy on tDNAs.

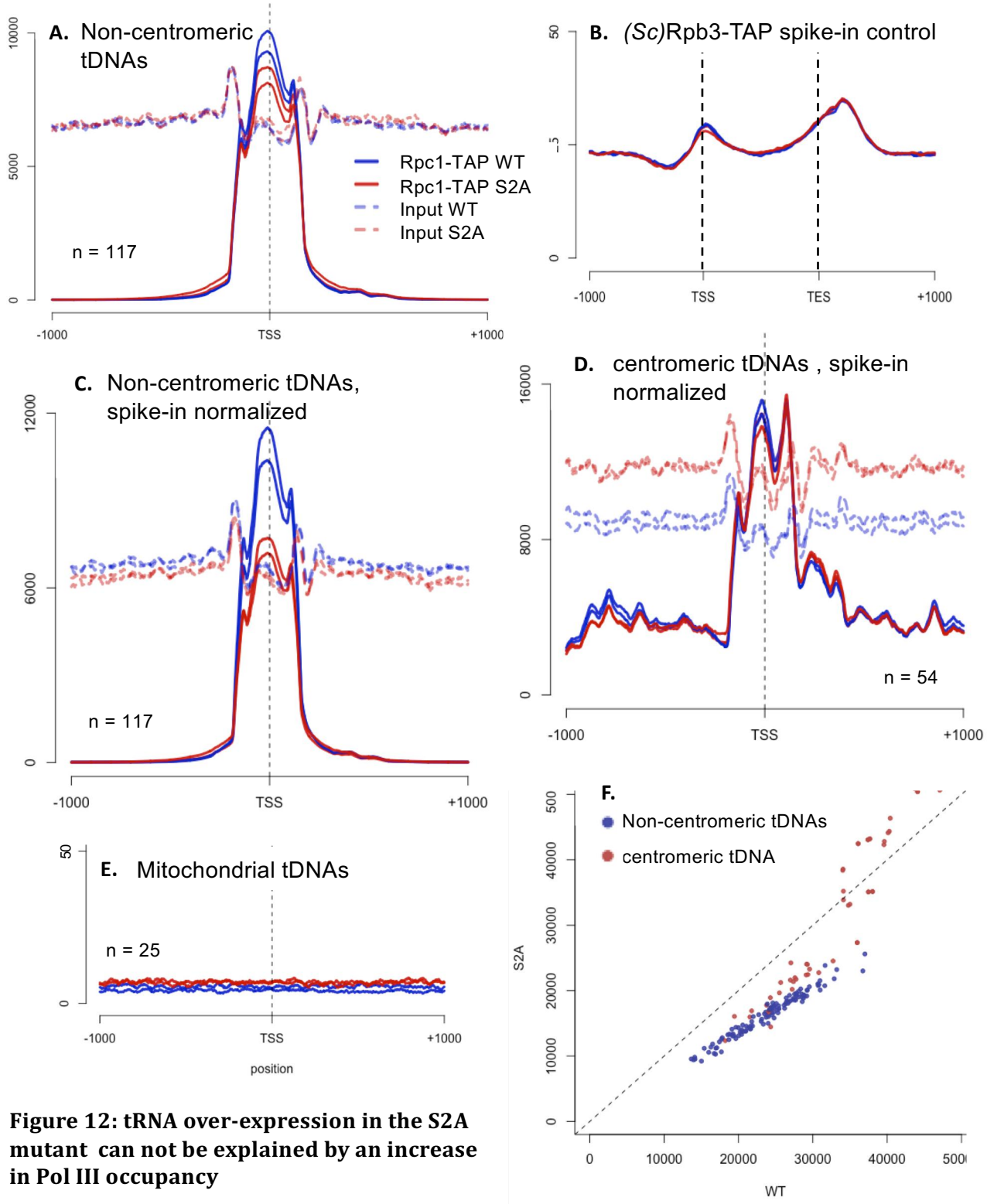


Figure 12: tRNA over-expression in the S2A mutant can not be explained by an increase in Pol III occupancy

A,C-E. Meta-gene average profile of ChIP-seq or input signal for Pol III (Rpc1-TAP) centered on tDNAs TSS for the indicated subset of tDNAs **(A)** The signal was normalized on the total number of mapped reads by condition. (C-E) The signal was normalized on the total number of spike-in reads by condition.

B. Meta-gene average profile of the spike-in (*S. cerevisiae* Rpb3-TAP) signal mapped on *S. cerevisiae* genome and scaled to annotated Pol II genes normalized on the total number of spike-in reads by condition.

F. Scatter plot of the Pol III peak intensity on tDNAs in the WT and S2A conditions summed over two replicates and normalized on the total number of spike-in reads by condition.

(3) Finally, it should be noted that our premise (the idea that the MNase and histone ChIP data are inconsistent) might not hold. Indeed, in a response to (Chereji et al., 2017), researchers from David Shore's lab argue against the idea that the MNase-sensitive particles are due to TF binding, based upon functional experiments where TFs are rapidly depleted (Kubik et al., 2017). They propose instead that histone ChIP assays have a limited capacity to capture these highly dynamic, MNase-sensitive "fragile" nucleosomes. Therefore, our failure to detect histone H3 changes in the S2A mutant does not necessarily imply that there is no change, an hypothesis that is difficult to experimentally validate. In any case, the nature of the MNase-sensitive particle observed at gene's promoters appears very controversial[6], and despite our thorough investigation, the – possibly related – peak of MNase resistance in the S2A mutant remains of unknown origin.

iii. tRNA over-expression in the S2A mutant cannot be explained by an increase in Pol III occupancy.

In budding yeast and higher eukaryotes, active transcription of tRNAs is globally correlated with Pol III occupancy on tDNAs chromatin (Orioli et al., 2016; Roberts et al., 2003). Therefore, we wondered whether the increase in (pre-)tRNAs transcripts – and also the increased MNase resistance – observed in the S2A mutant is accompanied with an increase in Pol III occupancy. To address this question at a genome-wide scale, we designed a ChIP-seq experiment using an epitope TAP-tagged Rpc1 (the biggest Pol III subunit). Classical ChIP-seq normalization methods, like scaling on total read counts, rely on the assumption that there is no global change in protein binding. As we cannot reasonably assume this in our conditions (all tRNAs are up-regulated after all), we set-up a spike-in-based normalization strategy for which we mixed 10% of a Rpb3-TAP *S. cerevisiae* culture to our samples at the beginning of the ChIP experiment, right before the cross-linking step. After sequencing, we scaled the reads mapped to *S. pombe* genome by the number of reads mapped on *S. cerevisiae* genome in order to control for both technical variation in sample preparation (chromatin extraction, IP efficiency, …) and for global changes in *S. pombe* Rpc1 occupancy.

Validating the Rpb3-TAP *S. cerevisiae* spike-in, the average profile on *S. cerevisiae* Pol II genes display the expected trend for a Pol II subunit (Mayer et al., 2010), peaking at the beginning and end of genes regardless of the condition (**figure 12.B**). This confirms that Rpb3 was properly immuno-precipitated in all of our samples.

---

[6] To cite David Shore and colleagues: *"We currently do not know whether the dynamic nature of [fragile nucleosomes] is due to more frequent unwrapping of DNA from the histone octamer core, rapid translational motion of the DNA over the histone core, direct destabilization by bound [general regulatory factors] and/or chromatin remodelers, a sub-octamer core composition, or some combination of these factors, but this is a question of considerable interest."* (Kubik et al., 2015)

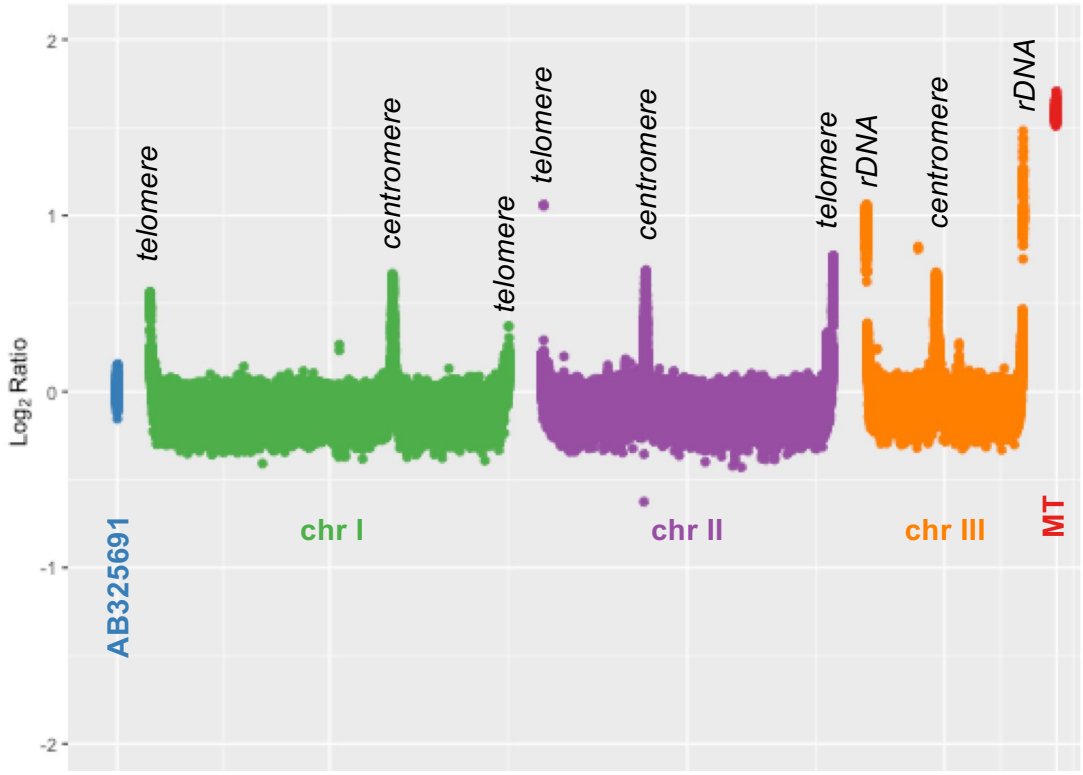


Figure 13: copy number variation or chromatin accessibility change in the S2A mutant.

A. Log2-ratio between S2A and WT chromatin input. The figure and ratios were calculated using CNV-seq.

On *S. pombe* tRNA genes, Rpc1-TAP levels were found highly enriched compared to background (**figure 12.A**). Confirming the specificity of the Rpc1 immunoprecipitation, no signal was detected at mitochondrial tRNA genes that are transcribed by a dedicated RNA polymerase (**figure 12.E**). Comparison of the Rpc1-TAP signal in the S2A and wild-type conditions on nuclear tRNA genes reveals a slightly lower level in the S2A mutant (**figure 12.A**); a trend that is accentuated after spike-in normalization (**figure 12.C**). This surprising result indicates that the accumulation of pre-tRNAs in the S2A mutant is not due to an increase in Pol III occupancy at tDNAs. Instead, it could be that Pol III transcriptional or recycling rate is increased in the S2A mutant – independently of Pol III occupancy – or that the tRNA misregulation occurs at a post-transcriptional level, for instance through the stabilization of the tRNA precursors. That last hypothesis could be validated using assays directly measuring transcription rate, rather than steady-state RNA level of bulk polymerase occupancy, for instance 4-tU labelling-based methods {Mata, 2017 #373}.

Strikingly, Rpc1 occupancy on a subset of tRNA does not follow the same trend and is either unaffected or higher in the S2A mutant (**figure 12.F**). Those genes are highly occupied in both conditions and are clustered within the centromeric regions, although some centromeric tDNAs (usually those located at the border of the centromeres) behave like the other tDNAs scattered throughout the genome. Accordingly, the average profiles for Rpc1-TAP at centromeric tDNAs, in contrast with the non-centromeric ones, are not very different between the WT and S2A conditions. However, this distinction between the two sets of tRNA genes appears artefactual as the average profiles for the input fractions at centromeric tDNAs are significantly higher in the S2A mutant (**figure 12.D**), suggesting either copy number variation (CNV) or differential chromatin accessibility at centromeres.

iv. CNV and chromatin accessibility changes in the S2A mutant

In order to systematically assess changes in chromatin accessibility or CNV genome-wide in the S2A mutant, we applied the sliding windows approach implemented in CNV-seq (Xie and Tammi, 2009) to our data, comparing the number of mapped reads in the input for every Kb of *S. pombe* genome. Systematically, more reads are mapped at sub-telomeric and centromeric loci of each of the three chromosomes in the S2A mutant (**figure 13**), indicating that those regions are either in multiple copies or present chromatin alterations that favors their recovery during sample preparation. Intriguingly, it was recently reported that Ser2-P in fission yeast is required for pericentromeric (Kajitani et al., 2017) and subtelomeric (Inada et al., 2016) silencing, although the mechanisms at play are still controversial: one study claims that Ser2-P is required for heterochromatin formation at subtelomeres based on H3K9me2 ChIP-microarray data (Inada et al., 2016), while the other concludes that the heterochromatin structure is almost completely retained in the S2A mutant at pericentromeres, based on ChIP-qPCR data on multiple heterochromatin factors (although they also observed a decrease in H3K9me2) (Kajitani et al., 2017).

Regardless of the mechanism, those studies suggest that the changes in the input level are caused by changes in the chromatin structure at centromeres and subtelomeres. Therefore, caution should be taken when interpreting differences in ChIP-seq

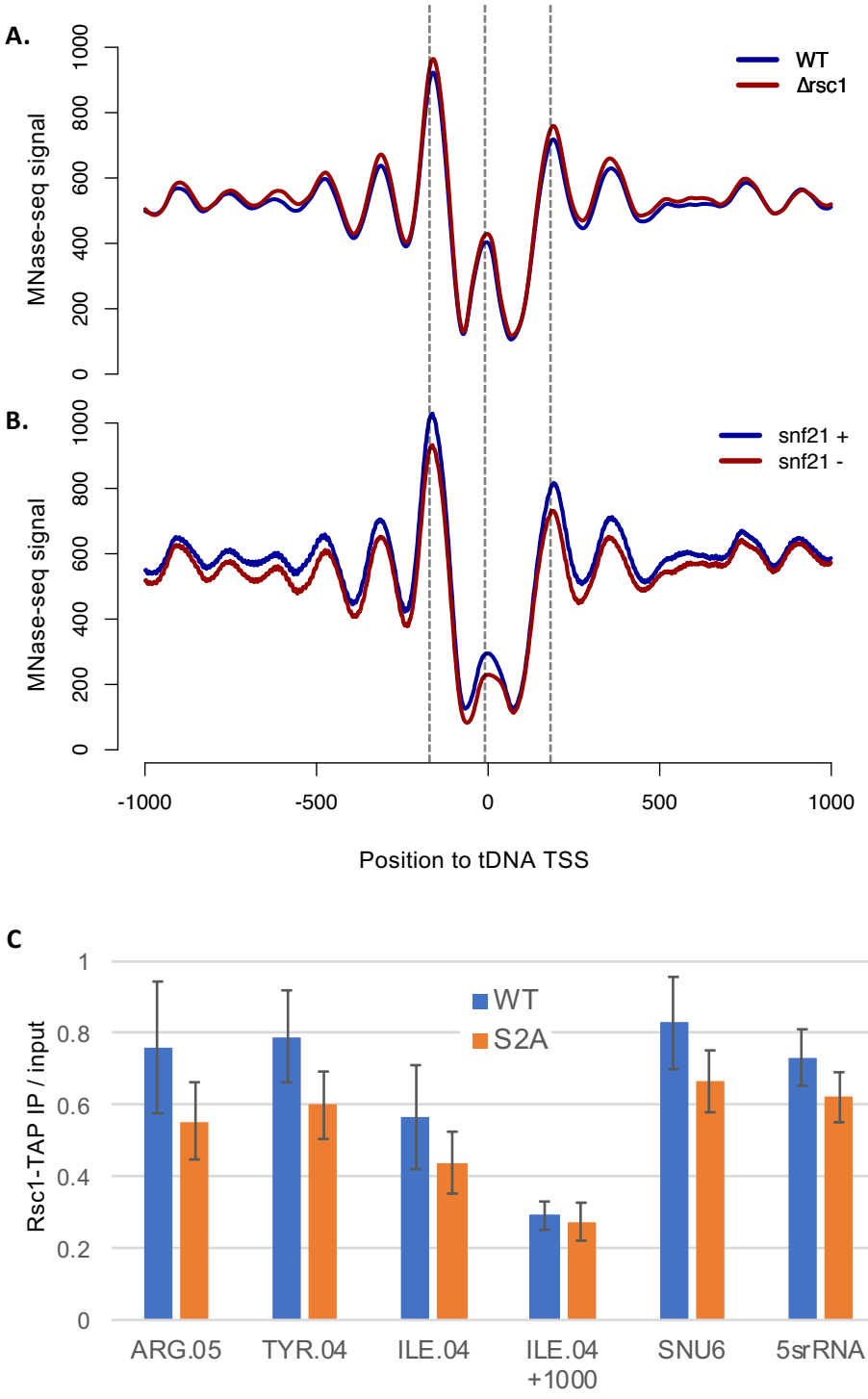


Figure 14: The RSC complex is not involved in Ser2-P dependent chromatin regulation at tRNA genes.

A. Meta-gene average profile of MNase-seq signal in the indicated conditions. Data from (Materne, 2016).

B. Meta-gene average profile of MNase-seq signal in the indicated conditions. The expression of *snf21*, the gene encoding the catalytic RSC subunit, is under the control of a tetO promoter that is activated (*snf21 +)* or repressed *(snf21 -).* Data from (Yague-Sanz, 2017).

C. ChIP-qPCR experiment on the TAP-tagged RSC subunit Rsc1. Values are expressed as the percentage of IP signal over the input, and the error bars represent the standard error of the mean over four biological replicates.

experiments in the S2A mutant. In our case, the difference observed in Rpc1 level at centromeres tDNAs are indeed artefactual, as after scaling to the input, all tDNAs behaved similarly, i.e, they were all less occupied by Pol III in the S2A mutant (data not shown).

Finally, we observed a 2-3-fold increase in the number of rDNA repeats at the ends of chromosome III (unless this difference is also due to a change in chromatin accessibility) and about three times more mitochondrial DNA in the S2A mutant (**figure 13**). How Pol II Ser2-P is connected to mitochondria biogenesis and rDNA copy number control is currently unknown and could become the focus of future studies.

v. The RSC complex is not involved in the Ser2-P-dependent chromatin regulation at tDNAs.

Work in budding yeast highlighted a critical role for the RSC (Remodel the Structure of Chromatin) – a chromatin remodeler complex present at all Pol III genes (Ng et al., 2002; Parnell et al., 2015) – in establishing nucleosome-depleted regions at tRNA genes. Specifically, inactivation of the RSC catalytic subunit leads to a general reduction of *de novo* Pol III transcription and a bulk increase in nucleosome density at tRNA genes (Kumar and Bhargava, 2013; Mahapatra et al., 2011; Parnell et al., 2008). Because this increase is reminiscent of the changes in MNase resistance we observed in the S2A mutant (although the impact on Pol III transcription is not), we wondered whether Pol II Ser2-P could be required for the RSC complex recruitment and/or activity at fission yeast tDNAs. Indeed previous works from our group revealed that Ser2-P, through the taming of the Set1 histone methyl-transferase recruitment, is important for histone acetylation and RSC recruitment at a subset of Pol II gene promoters (Materne et al., 2015; Materne et al., 2016).

In order to assess the role of the RSC complex at tRNA genes in fission yeast, we re-analyzed MNase-seq data in conditions where the function of the RSC complex is impaired. By deleting *rsc1*, a non-essential subunit (Materne et al., 2016), we could not observe differences in MNase resistance at tDNAs (**figure 14.A**). Similarly, by switching-off with a Tet-Off system[7] the expression of the gene encoding the essential catalytic subunit of the RSC complex, *snf21* (Yague-Sanz et al., 2017), no difference was observed (**figure 14.B**). Those results argue against the conservation of the role of the RSC complex in excluding nucleosomes at tDNAs in fission yeast. However, as thoroughly discussed the **Appendix 1**, the activity of the RSC complex is not fully abolished in those strains and we cannot exclude that this residual RSC activity is sufficient to generate nucleosome depleted regions at tDNAs.

In relation to the previous point that interrogates the nature of the peak of MNase resistance at tRNA genes, we cannot help but notice that the intensity of this peaks vary

[7] In a Tet-Off system, the expression of a gene is switched off rapidly upon the addition of tetracycline. Tetracycline binds and inhibits a transactivator protein that, when free, activates the expression of a gene of interest through the binding of tetO operator sequences placed upstream of the gene promoter (Gossen & Bujard, 1992)

from experiments to experiments, even for the exact same strains, processed by the same lab (although at different times) and analyzed using the same bioinformatics pipeline: very low for the WT from (Materne et al., 2015) (**figure 10.B**, higher panel) but significantly higher in the WT from (Materne et al., 2016) (**figure 14.A**). This variation reinforces the idea that those sites are "fragile" – histone or non-histone – sites of MNase resistance, that are very sensitive to technical variation in the MNase experiments and whose nature is still largely under debate.

Finally, to assess the putative role of Ser2-P in recruiting the RSC complex at tRNA genes, we immunoprecipitated an epitope tagged version of one of its subunits, Rsc1. Although we observe a small decrease of RSC occupancy at the tDNAs tested – and not at a control intergenic region –, the difference is not statistically significant, arguing against a major role of Ser2-P in recruiting the RSC remodeling complex at tRNA genes. This is consistent with the literature that describes a direct interaction between Rbp5 (a subunit shared between Pol I, Pol II and Pol III) and the Rsc4 RSC subunit (Soutourina et al., 2006).

vi. Final comments

As a whole, this chapter on chromatin changes in the S2A mutant reports mixed results. Although the chromatin structure appears altered at tRNA genes (and other loci) in the S2A mutant, it is still unclear in what way it is altered. Our negative results point that it is probably not due to differential TF binding, RSC binding, Pol III binding or nucleosome binding. As a consequence, it is currently unclear how this chromatin alteration could affect Pol III transcription and whether it is related at all to the increased level of pre-tRNAs observed in the S2A mutant.

To further investigate this issue, one could test the differential binding of other protein candidates, such as condensin. Indeed, condensin was reported to bind tRNA genes and to bridge them together, allowing centromeric localization of dispersed tRNA genes, which negatively correlates with tRNA expression (Iwasaki et al., 2010). Potentially altered nuclear localization of tRNA genes in the S2A mutant could impact tRNA metabolism in multiple ways, such as transcription (Iwasaki et al., 2010) and export (Chen and Gartenberg, 2014), providing a possible mechanistic basis for the tRNA accumulation in the S2A mutant.

F. The exonuclease Rrp6 targets pre-tRNAs for degradation

i. rrp6 *deletion suppresses the growth defect of* rpc25-flag

In the previous analyses, we found little evidence that the accumulation of pre-tRNAs in the S2A mutant is directly caused by a modulation of Pol III transcription. Indeed, while Pol III occupancy is known to generally correlate with its activity (Roberts et al., 2003), Pol III ChIP-seq revealed that there is actually less Pol III at tRNA genes in the S2A mutant. Moreover, deletion of *maf1,* the gene encoding for the only known global Pol III repressor in yeast, barely affected the expression of the tRNA tested, suggesting that Pol III is largely unrepressed in our conditions where nutrients are fully available. In that unrepressed setting, it is difficult to imagine how the loss of Ser2-P could boost tRNA transcription directly. Therefore, we wondered whether Ser2-P could act at a post-transcriptional level by being involved in pre-tRNA degradation.

The exosome contains two catalytic subunits, the essential endo-/exo-nuclease Dis3, and the non-essential 3'-5'exonuclease Rrp6. Interestingly, deletion of *rrp6* within the *rpc25-flag* strain almost completely restored wild-type-like growth in the strain, reminiscent of the suppression by the S2A mutation (**figure 15.B**). This also suggests that, in fission yeast, just like in budding yeast, at least some of the pre-tRNA degraded by the exosome can become fully matured, functional tRNAs, and as such, probably lack processing defects.

ii. Rrp6 deletion causes pre-tRNA accumulation

At the RNA-level, the *Δrrp6* deletion causes pre-tRNAs to accumulate massively, in a way that (over-)compensate the decrease in pre-tRNAs of the *rpc25-flag* strain (**figures 15.A, 15.C**). However, quantification of the tRNA precursors and mature tRNAs show that the various isoforms are differentially affected (**figure 15.D, right panel**). Full length pre-tRNAs (*pre-*) accumulate the most after *rrp6* deletion; between 4- and 8- folds over WT depending on the tRNA tested. In contrast, the 5'-processed tRNA precursors (*5'*) reach levels lower than the WT while the 5'-3'-processed tRNA precursors (5'-3') mildly accumulate (between 1- and 2-folds over WT). Finally, mature tRNAs accumulate about 2-folds over WT.

This suggests that *rrp6* preferentially targets for degradation full length pre-tRNAs although more insight into the complex dynamics of the various tRNA processing steps involved would be required to conclude strongly on these results. Nevertheless, it is intriguing to see that the S2A mutant displays almost the same patterns at a smaller scale (**figure 15.D, left panel**): Full length pre-tRNAs accumulate the most (between 2- and 3-folds), 5'-processed precursors reach a slightly lower-than-WT level and mature tRNAs accumulate between 1- and 2- folds. Those similarities could indicate that the S2A mutation affects tRNA expression in a way similar to the *rrp6* deletion if it was not for the effect on the 5'-3'-processed tRNA precursors. Indeed, these precursors accumulate in the S2A mutant almost to the same level as the full-length precursors.

Accumulation of unspliced 5'-3'-processed is typical of strains defective in nuclear tRNA export – a prerequisite for pre-tRNA splicing –, such as strains deleted for the

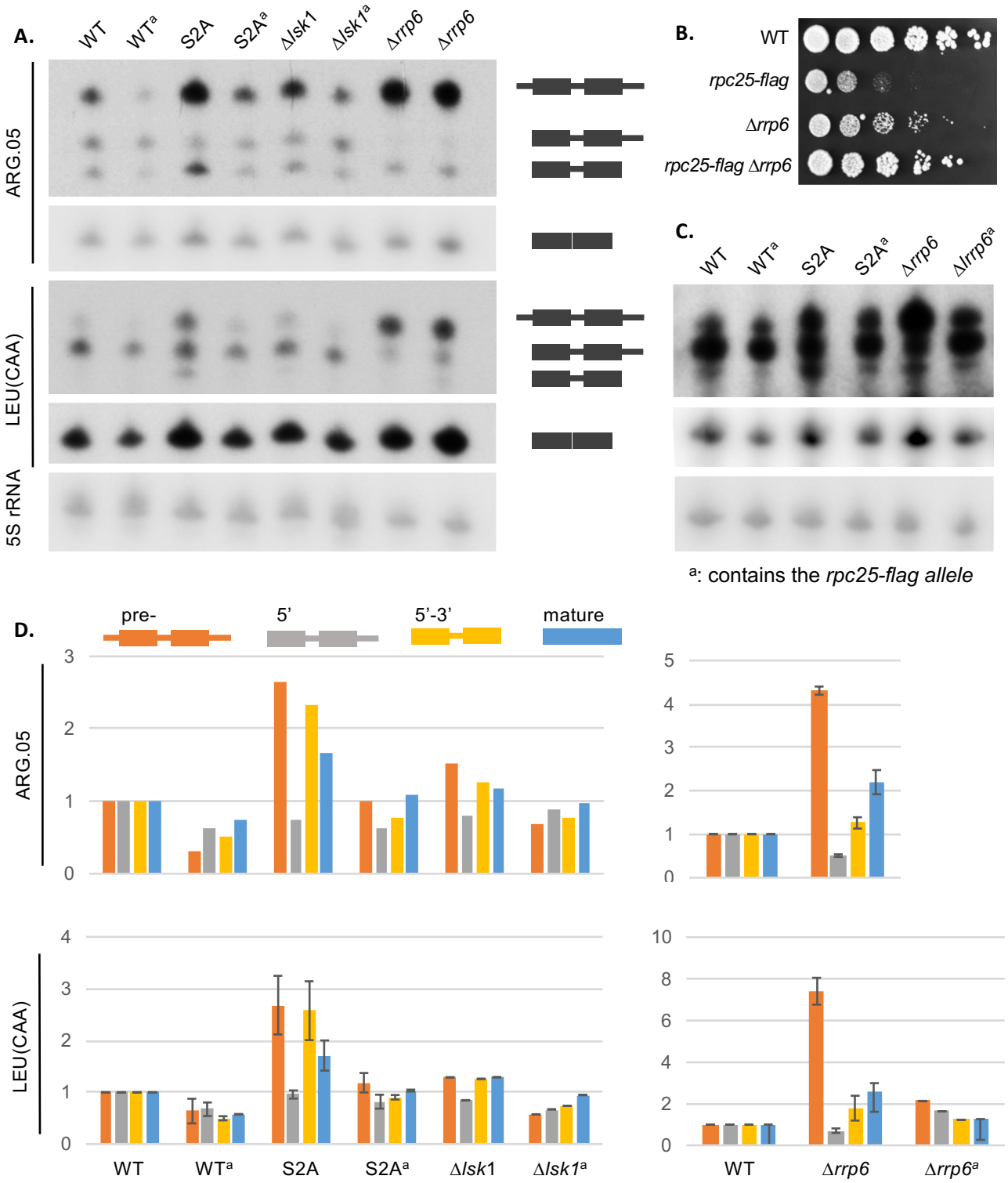


Figure 15: The exonuclease Rrp6 targets pre-tRNAs for degradation.

A,C. tRNA northern blot targeted against SPBTRNAARG.05, LEU(CAA) or 5S rRNA in the indicated strains.
[a]: contains the *rpc25-flag allele*

B. Growth assay of the indicated strains. Precultures were grown overnight then diluted in liquid YES media to OD 0.2 and incubated with agitation at 32°C until OD 0.5. From there, 5-fold dilutions were spotted on YES-agar plates and incubated during 3 days at 32°C.

D. Quantification of the pre- and mature tRNA from (A) and (C). After background subtraction, tRNA levels were normalized to the mature 5S rRNA level.

conserved *los1* nuclear tRNA exportin (Chatterjee et al., 2017; Cherkasova et al., 2011). However, it is difficult to incriminate a processing defect for the accumulation of pre-tRNAs in the S2A mutant, as mature tRNAs accumulate as well and rescue the transcriptional defect of the *rpc25-flag* allele.

iii. lsk1 *deletion causes a mild pre-tRNA accumulation, yet fully suppresses* rpc25-flag

From the same experiment, we also observe that deletion of Lsk1, the main Ser2-P kinase in *S. pombe* has a similar impact on (pre-)tRNA levels as in the S2A mutant although the changes compared to the WT are of much smaller amplitude (**figure 15.D, left panel**). This discrepancy could be explained by residual Ser2-P observed in the *Δlsk1* mutant (data not shown). Remarkably, while the level of tRNA precursors cannot be fully rescued in the Δ*lsk1 rpc25-flag* mutant, *lsk1* deletion still restores WT-like levels of mature tRNAs and – probably as a consequence – completely suppresses the growth defect of the *rpc25-flag* allele (**figure 3.C**).

iv. Two-way suppression for the Δrrp6 rpc25 *double mutant*

Surprisingly, while the double *Δrrp6 rpc25-flag* mutant is growing faster than the simple *rpc25-flag* mutant, it is also growing better than the single *Δrrp6* mutant (**figure 16.A**). This result suggests that one of the reasons for the *Δrrp6* growth defect is the accumulation of pre-tRNAs, an accumulation that would be tuned down by the hypomorphic *rpc25-flag* allele.

In addition, despite that the exosome is known to cooperate with the TRAMP complex to target RNAs for degradation (Schmidt and Butler, 2013), deletion the TRAMP subunit Cid14 (homologous to Trf5 in *S. cerevisiae*) does not recapitulate the suppression of *Δrrp6* strain. Moreover, deletion of the gene encoding the Xrn2 cytosolic exonuclease – responsible for the rapid tRNA decay (Xrn1 homolg) [REF]– is also not suppressive. These result highlight the specificity of the suppression of the hypomorphic *rpc25-flag* allele by the *Δrrp6* strain, but more work would be required to assess whether Rrp6 degrades pre-tRNA in an TRAMP-independent fashion, or even in an core exosome-independent fashion (Graham et al., 2009).

v. Is Rrp6 connected to Ser2-P?

To assess whether the increase in pre-tRNA observed in the S2A mutant would be dependent on Rrp6 degradation, we tried to generate a *Δrrp6* S2A double mutant. However, despite our best efforts, we never could obtain this strain, be it by crosses followed by random spore analysis or by transformation targeting *rrp6* deletion in the S2A background. In the case of difficult crosses, tetrad dissection is the method of choice as it allows to see whether a particular combination of alleles is lethal. In the S2A mutant however, tetrad dissection is impractical given its penetrant sterility. In consequence, while our inability to obtain it make it likely that the S2A *Δrrp6* double mutant is lethal, we could not prove it.

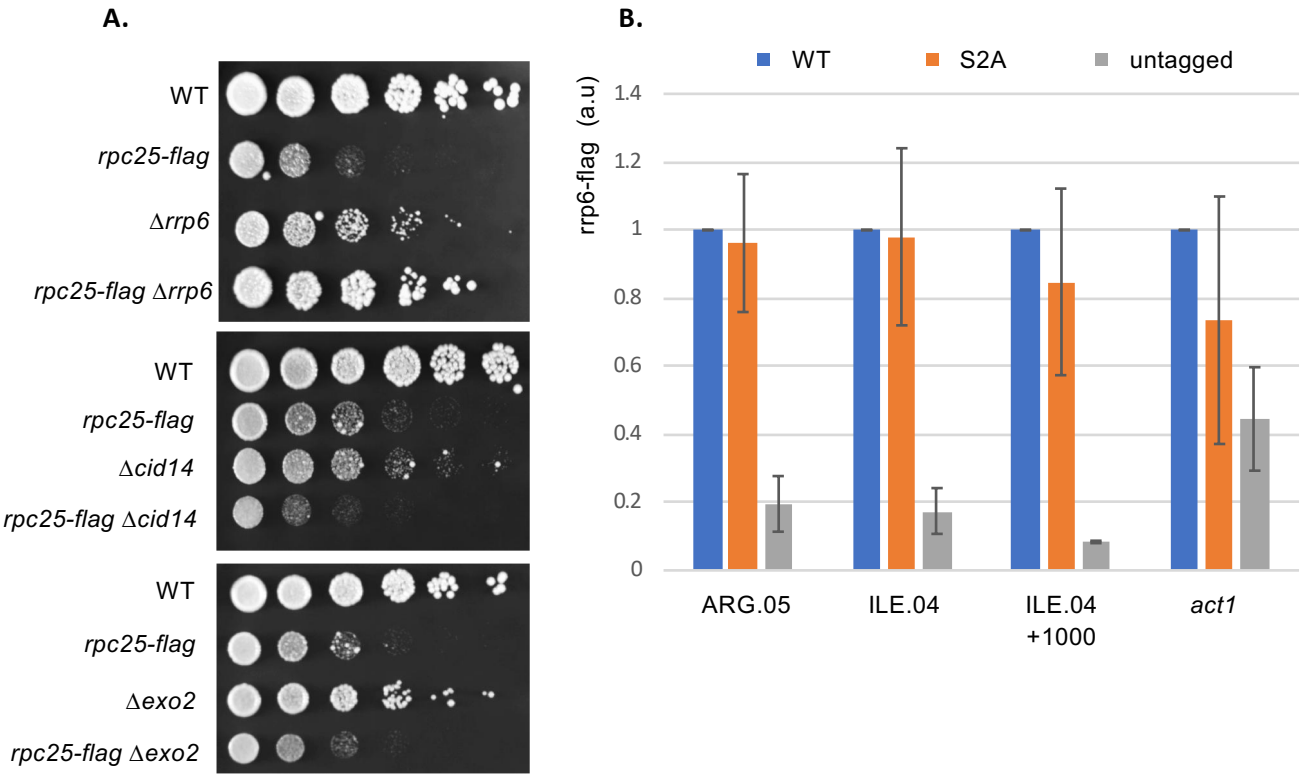


Figure 16: Rrp6 suppression of *rpc25-flag* is TRAMP-independent but Rrp6 does not differentially localize at tRNA genes..

A. Growth assay of the indicated strains. Precultures were grown overnight then diluted in liquid YES media to OD 0.2 and incubated with agitation at 32°C until OD 0.5. From there, 5-fold dilutions were spotted on YES-agar plates and incubated during 3 days at 32°C.

B. ChIP-qPCR experiment on the TAP-tagged TFIIIB subunit Brf1. Values are expressed as the percentage of IP over the input, normalized on the WT level and the error bars represent the standard error of the mean over two biological replicates.

As an alternative approach to study the connection between Ser2-P and Rrp6, we used an epitope-tagged *rrp6* subunit to assess Rrp6 occupancy on tRNA genes in the S2A mutant. Indeed, the nuclear exosome has been reported as involved in co-transcriptional processing and as such, can interact with chromatin (de Almeida et al., 2010; Hessle et al., 2012; Lemay et al., 2014). Accordingly, we were able to immunoprecipitate the exosome at the two tRNA genes tested to a level of at least 5-fold above background (untagged strain). In contrast, at the *act1* control loci, only a 2-fold increase over background was observed, indicating that the exosome is preferentially found at tRNA genes (**figure 16.B**). Arguing against a role for Ser2-P in recruiting Rrp6, no significant difference of enrichment was observed in the S2A mutant compared to the wild-type strain.

In conclusion, despite troubling similarities in how pre-tRNAs react to both the loss Ser2-P and the loss of Rrp6, we were not able to demonstrate a connection between the two. To provide further insight into the question, we are currently assessing pre-tRNA stability in the S2A mutant in order to unambiguously distinguish a transcriptional effect from a stability effect.

G. The SAGA connection

i. Isolation of Rpc25-flag *suppressors*

Despite our candidate-based assays exploring connections with chromatin changes and exosome degradation, the mechanistical connection between Ser2-P and pre-tRNA remains largely elusive. Therefore, we wanted to try to unravel this connection using a more unbiased approach, taking advantage of the *rpc25-flag* phenotype. The idea for the screen stem from the observation that some fast-growing *rpc25-flag* colonies, visible in most growth assays (see figure 16.A, lower panel, for instance), appear to bypass the growth defect imposed by the *rpc25-flag* allele. Those naturally occurring suppressors where intriguing, as they mimicked the loss of Ser2-P in the *rpc25-flag* strain. Therefore, we set up a screen aimed at the identification of the suppressive mutation in order to unravel potentially new regulatory layers for Pol III transcription, which might provide the missing link between Ser2-P and tRNA expression.

Eleven natural suppressors were isolated from growth at restrictive temperature (32°C or 37°C). In order to avoid working with trivial second-sites suppressors, we first screened by PCR, western-blot and/or sanger sequencing for the integrity of the *rpc25-flag* allele. Five of the suppressors had partially of totally lost their *flag* fusion at the end of *rpc25*, incidentally confirming that the growth defect of the *rpc25-flag* strain is due to the presence of the *flag*.

The remaining six suppressors, named s1 to s6 restored wild-type-like growth at restrictive temperature (**figure 17**), with the exception of s1 that is still temperature sensitive at 37°C. This result makes sense as s1 is the only suppressor isolated at 32°C where the selective pressure is different, and likely less stringent, than at 37°C (**figure 5.C**).

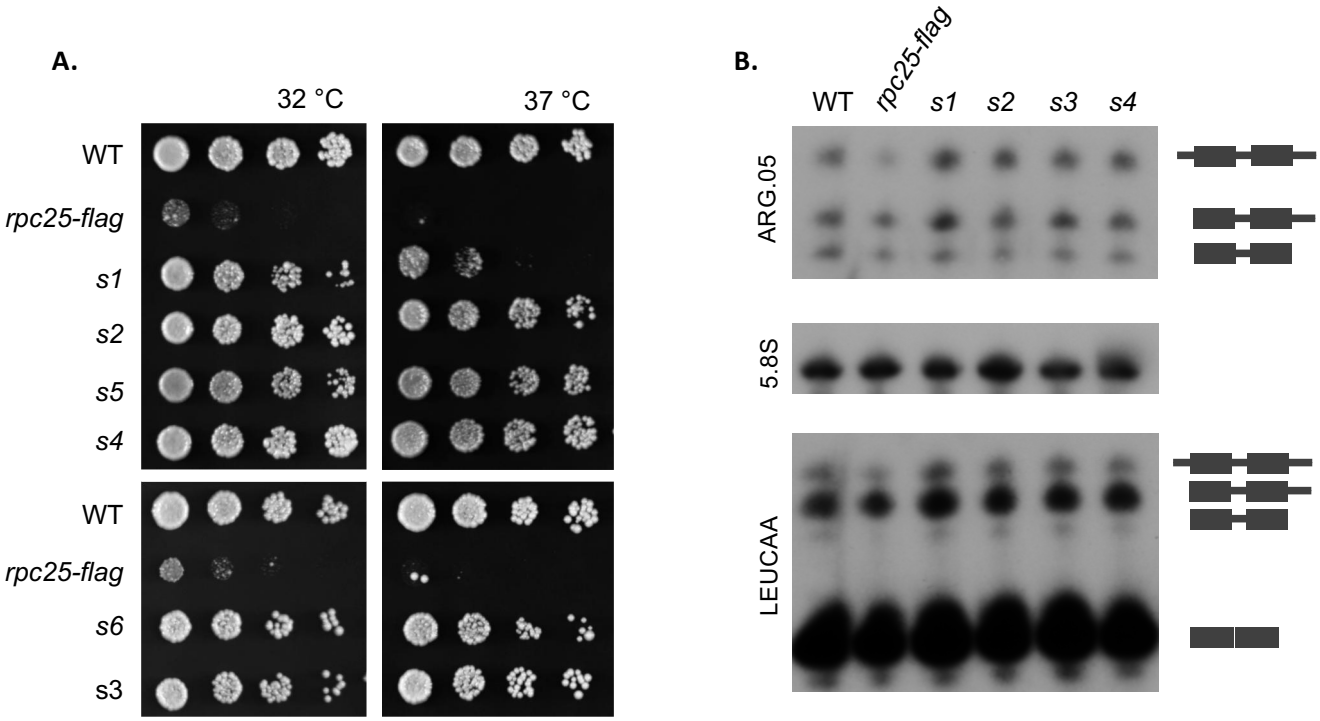


Figure 17: Growth assay the naturally isolated *rpc25-flag* suppressors.

A. Growth assay of the indicated strains. Precultures were grown overnight then diluted in liquid YES media to OD 0.2 and incubated with agitation at 32°C until OD 0.5. From there, 5-fold dilutions were spotted on YES-agar plates and incubated during 3 days at 32°C.

B. tRNA northern blot targeted against SPBTRNAARG.05, LEUCAA and 5.8S rRNA in the indicated strains.

Table 4: Identification of the suppressive mutations.

suppressor	T° of selection	genotype[a]	effect	% of ORF affected[b]
s1	32°C	*rpc25-flag pcr1_18+T*	frameshift	96.5
s2	37°C	*rpc25-flag tra1_3767+A*	frameshift	66.1
s3	37°C	*rpc25-flag sgf73_967+A*	frameshift	13.5
s4	37°C	*rpc25-flag tms1_G155T*	Gly to Val	NA
s5	37°C	*rpc25-flag mug133_C500T*	silent	NA
s6	37°C	*rpc25-flag tra1_G3145T*	STOP	71.7

[a]: for the suppressive mutations, the naming convention follow this scheme: '*pcr1_18+T*' means that a T is inserted 18 nucleotides after the annotated start codon of the *pcr1* gene model. '*tms1_G155T*' means that the G positioned 155 nucleotides downstream the annotated start codon of the *tms1* gene model is replaced by a T.

[b]: proportion of the open reading frame downstream the frame shift or stop codon.

In the golden age of genetics, the suppressive mutations could have been identified using tedious genetic mapping but (thankfully) the advent of high throughput sequencing techniques considerably facilitated this step. Indeed, through the whole genome sequencing of the suppressors, we easily sequenced and identified the mutated alleles by comparison with the original *rpc25-flag* strain using VarScan (Koboldt et al., 2012). Only one indel or single-nucleotide variation by suppressor was detected in ORFs with high confidence, that is, with at least 75% of reads supporting that variation (**table 4**). Although we cannot exclude to have missed some variation (in particular, the mode of sequencing chosen – single-end 50 nt – does not allow to unveil complex rearrangements), we do not expect to have more than one or two mutations per suppressor given that the yeast mutated at their natural rate (about 2.00×10^{-10} mutations per site per generation (Farlow et al., 2015)) during only a few days.

ii. Mutations within genes encoding subunits of the SAGA complex suppress a defect in Pol III transcription.

Strikingly, three out of the six suppressors identified (frameshift in *sgf73* and frameshift and stop codon in *tra1*) had disruptive mutations within subunits of the Spt-Ada-Gcn5 acetyltransferase complex (SAGA). SAGA is an evolutionarily conserved transcriptional coactivator of Pol II comprising 18 to 20 subunits organized within five (six in metazoan) functional modules with distinct regulatory activities, reviewed in (Helmlinger and Tora, 2017; Rodríguez-Navarro, 2009):

- Interaction with transcription activators (by **Tra1**)
- histone acetyl transferase activity (by Gcn5 acetyl-transferase)
- deubiquitinylation(DUB) (by the Ubp8 deubiquitinase – and **Sgf73**)
- recruitment of the TATA-box binding protein (by Spt3 and Spt8)
- binding of methylated histone (by Sgf29)

When localized at gene promoters, SAGA is thought to activate gene expression through the recruitment of the TATA binding protein and through histone acetylation and deubiquitinylation. While a seminal study estimates at ~10% the number of Pol II gene

promoters controlled by SAGA (Huisinga and Pugh, 2004), a more recent study found that the expression of virtually all Pol II genes is SAGA-dependent (Baptista et al., 2017). To our knowledge, no role in controlling tRNA expression has been reported for the SAGA complex, yet it is found at tRNA genes in drosophila (Weake and Workman, 2012). Adding to the complexity of understanding SAGA roles, deletion of individual SAGA subunit lead to gene-specific expression changes (Helmlinger et al., 2011; Lenstra et al., 2011).

In our screen, two SAGA subunits from two different modules came out of our screen, Tra1 and Sgf73. The Tra1 SAGA subunit, an unusually large protein (422 Kda), is not essential for viability in *S. pombe* in contrast with its homologs in budding yeast and mammals (Calonge et al., 2010). Intriguingly, the Tra1 homolog TRRAP is known to be recruited at Pol III transcription sites, along the Gcn5 histone acetyltransferase (probably within the SAGA complex), by the proto-oncogene c-Myc to promote tRNA expression in human cell lines (Kenneth et al., 2007). However, this link between Tra1 and Pol III expression is difficult to connect with the suppression of the *rpc25-flag* growth defect by Tra1 disruption. Indeed, c-Myc is not conserved in yeast. In addition, the disruption of a factor promoting tRNA expression would probably not compensate a hypomorphic Pol III allele.

In fission yeast, Tra1 is required for the expression of only a subset of the SAGA-dependent genes. Intriguingly, although no global SAGA assembly defect was detected in mutants lacking Tra1, components of the DUB module such as Ubp8 and Sgf73 were underrepresented in the proteins associated with the SAGA subunit Ada2 in a *Δtra1* strain (Helmlinger et al., 2011). Therefore, an alternative hypothesis to the idea that Tra1 disruption suppresses a Pol III defect directly would be that the suppression occurs through a less efficient recruitment of the DUB module within SAGA, including Sgf73, the other SAGA subunit uncovered in our screen.

The regulation of histone H2B ubiquitination (H2Bubi) has important, yet complex, implications in the control of gene expression. Indeed, H2Bubi affects transcription differentially depending on where it is deposited: it acts as a repressor at gene promoters, yet, it facilitates elongation through favored nucleosome assembly/disassembly (Batta et al., 2011). This dual role has profound implication in processes as important as the control of sexual differentiation in fission yeast, for which we showed that H2Bubi opposes RSC remodeling at the *ste11* promoter (Materne et al., 2016). In addition, mutations that abolish the DUB activity induce a decrease in Ctk1 (a Ser2-P kinase, homolog to Lsk1 in *S. cerevisiae*) recruitment, further connecting deubiquitination with Ser2-P (Wyce et al., 2007).

In budding yeast, Sgf73 is also implicated in ribosomal protein (RP) expression as the deletion of *sgf73* causes reduced expression of RP genes and replicative lifespan extension (Mason et al., 2017). The situation appears similar in *S. pombe* (Helmlinger et al., 2011), suggesting that the role of SAGA in controlling RP expression is conserved within eukaryotes, or at least between those two distantly related yeast species. Given the importance of a coordinated regulation of the expression of the components of the protein synthesis machinery (rRNAs, RP genes and tRNAs) (Warner, 1999), a possible explanation for the suppression of the Pol III transcriptional defect of the *rpc25-flag* strain could depend on the decrease of expression of other components of the protein synthesis machinery, such as the SAGA-dependent RP genes. However, Northern blot analysis of the

suppressors revealed that they restored the expression of the pre-tRNAs, at least in proportion to the level of mature tRNAs, the total amount of RNA loaded and the expression of the 5.8S rRNA, a restoration that cannot be explained solely by the possible down-regulation of the RP genes in the suppressors (**figure 17.B**).

Strengthening the idea that the disruption of SAGA-regulated processes causes *rpc25-flag* suppression, a 4th suppressor, in which the *pcr1* gene is disrupted, is connected to SAGA. Indeed, Pcr1 is a transcription factor that cooperates with its heterodimer partner Atf1 to recruit SAGA at gene promoters (Sansó et al., 2011).

The remaining two suppressive mutations identified in our screen are more difficult to interpret. One of these is located on the gene coding for Tms1, a predicted hexitol dehydrogenase that, intriguingly, binds human p53 expressed in fission yeast (Wagner et al., 1993). The other one is a silent mutation on the *mug133* gene, which is coding for an uncharacterized protein.

In conclusion, with three out of six suppressive mutations falling into genes coding for SAGA subunits, our screen uncovered a strong genetic link between the SAGA Pol II coactivator complex and Pol III transcription. However, more work will be required to confirm these suppressions, and even more work to comprehend the mechanism behind, which might be linked to the Ser2-P-dependent regulation of tRNA expression proposed through this work.

H. ADDENDUM: A global Pol II transcriptional defect in the S2A mutant ?

As Ser2-P is involved in transcriptional termination through the recruitment of subunits of the 3'-processing complex (Harlen et al., 2016), we wondered whether Pol II distribution on the genome would be affect in the S2A mutant, and therefore, we performed ChIP-seq experiments targeted against Pol II. Intriguingly, global normalization revealed few differences in the meta-genes profiles centered on protein-coding genes between the wild-type and S2A mutant, with almost no Pol II extending beyond the gene boundaries (**figure 18.A, top-left panel**). This result argue against a global and essential role for Ser2-P in triggering transcriptional termination.

In contrast, on tRNA genes, global normalization revealed increased Pol II level in the S2A mutant (**figure 18.A, top-right panel**). This result could have been interpreted in a model where Pol II has a positive effect on Pol III transcription on tRNA genes while Ser2-P helps Pol II to disengage the DNA template at these loci. The loss of Ser2-P would therefore cause a local increase in Pol II occupancy that could, in turn, lead to increased tRNA expression.

However, when *S. cerevisiae* chromatin spike-in is used to provide an unbiased normalization method, it leads to a very different interpretation of the results. Indeed, after spike-in normalization, Pol II occupancy on both protein-coding genes and tRNA genes is massively decreased in the S2A mutant (**figure 18.A, lower panels**), suggesting a global Pol II defect in this strain.. Somehow confirming this idea, Western blot analysis of Rpb3 abundance, an Pol II subunit, show that the amount of Rpb3 is significantly

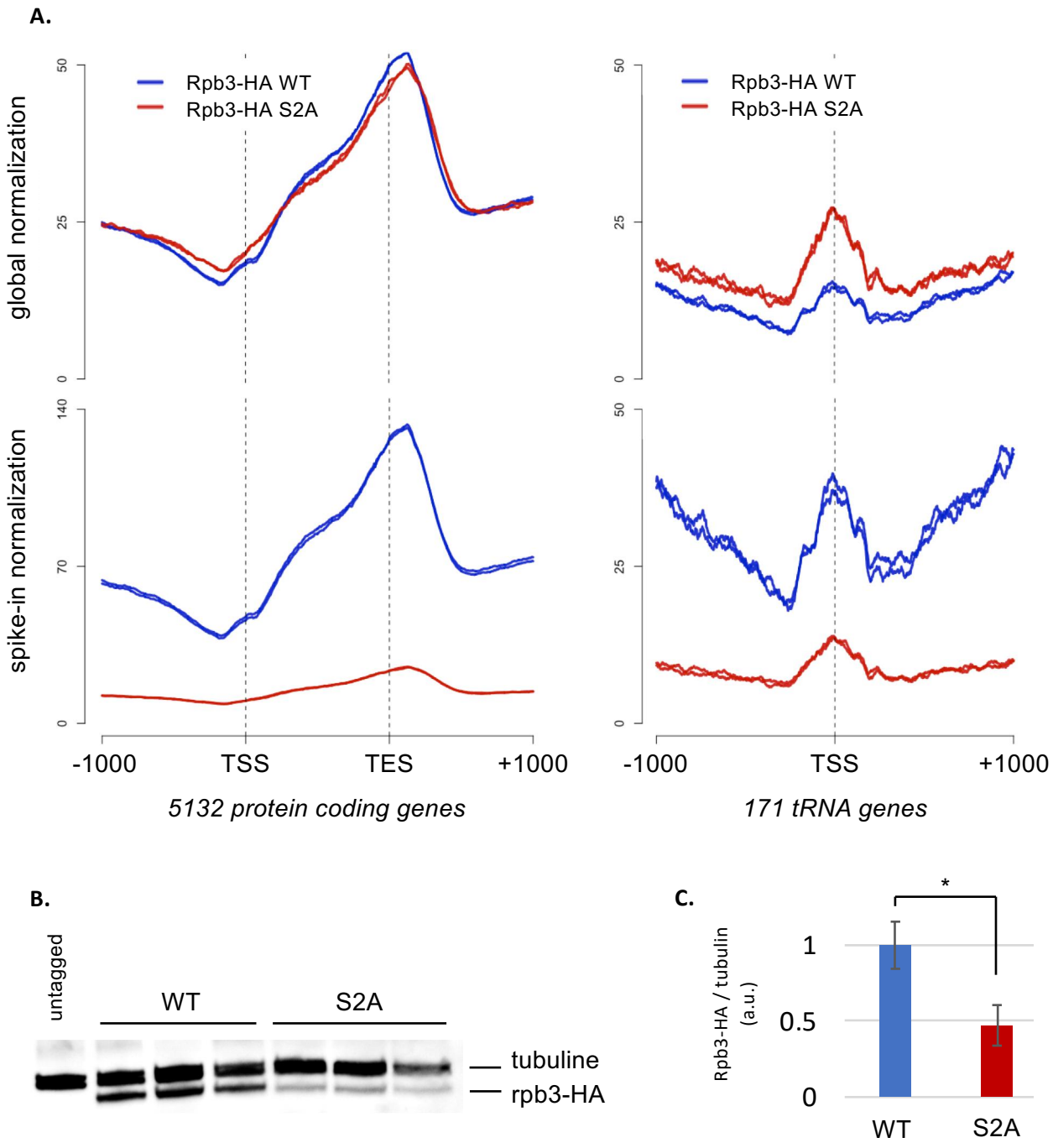


Figure 18: A global Pol II transcriptional defect in theS2A mutant ?

A. Meta-gene average profile of Pol II (Rpb3-HA) ChIP-seq signal (10% trimmed mean)**.** The profiles are either scaled on protein coding genes (left panels) or centered on tDNAs TSS (right panels). The signal was either normalized on the total number of mapped reads by condition (global normalization, top panels) or normalized on the total number of spike-in reads by condition (spike-in normalization, bottom panels). Two replicates were merged.

B. Western blot against Rpb3-HA. Tubuline serves as loading control.

C. Quantification of the Rpb3-HA signal from (A)**.** The signal was background-substracted and normalized over the tubuline in the imageQuant^{TL} software. The error bars represent the standard deviation from three biological replicates. T-test *pvalue* = 0.011.

decreased (more than two-folds) in the S2A mutant (**figure 18.B-C**). Together, those results suggest that an unknown mechanism (possibly proteolysis) causes the global quantity of Pol II subunits to decrease in the S2A mutant, and in consequence, the amount of chromatin-bound Pol II is also decreased.

How the cell can cope with such an heavy defect of Pol II transcription is currently unknown, yet as shown repeatedly throughout this manuscript, the S2A mutant has no growth defect in exponential growth. A possible explanation for this apparent contradiction would be "buffering" effects, *i.e.*, compensation of transcriptional defects through a global stabilizations of RNAs (Sun et al., 2013). This putative buffering effect could be asserted using methods to directly quantify the rate of transcription rather than the steady-state abundance of particular transcripts, for instance the 4-tU labelling method (Mata and Wise, 2017). If the buffering effect is genuine, then a plausible explanation for the accumulation of pre-tRNA in the S2A mutant would be that while mRNAs, transcribed at a lower rate, are stabilized by the global buffering effect to reach wild-type-like levels, tRNAs, transcribed at a normal rate, are also stabilized and reach higher levels. This model, however, fail to explain why Pol I-transcribed rRNA transcripts escape the buffering-dependent stabilization, as pre-tRNA accumulate compared to both rRNA (as shown repetitively by Northern blot within this thesis) or Pol II-transcribed RNAs (as seen by *RNA-seq*). In addition, the deletion of *exo2,* homolog of *xrn1*, the exonuclease responsible for the buffering effect in *S. cerevisiae (Sun et al., 2013)*, does not suppress the *rpc25-flag* allele (see figure 16.A, lower panel).

Altogether, while our results indicate a clear global Pol II transcriptional defect, a defect that could possibly be involved in the (still elusive) mechanism of tRNA over-expression in the S2A mutant, it is not sufficient to explain such mechanism. To disentangle long-term compensatory effects to direct effects associated with the loss of Ser2-P, one could take advantage of conditional mutants of the Ser2-P kinase, such as the *lsk1-as* mutant (Materne et al., 2015).

3. CONCLUSION

i. Global and gene-specific requirements for Ser2-P in fission yeast.

In a methodic assessment of changes in the transcriptional landscape in a S2A mutant, we uncovered global and gene-specific requirements for Ser2-P in fission yeast. As for the global aspect, the most prominent feature of the S2A mutant is a general increase in the level of the transcripts falling outside the genes boundaries, especially in anti-sense of the annotated genes. This increase is probably caused by an augmentation of Pol II cryptic transcription due to the loss of Ser2-P, involved in the recruitment of the H3K36 methyltransferase Set2. Set2 activity leads to the Rpd3 histone deacetylase recruitment, that ultimately helps to close the chromatin in the wake of Pol II transcription. This process has been extensively covered in *S. cerevisiae* studies (Carrozza et al., 2005; Keogh et al., 2005; Kim and Buratowski, 2009; Li et al., 2003; Li et al., 2002; Lickwar et al., 2009; McDaniel et al., 2017; Venkatesh et al., 2016; Xiao et al., 2003) and to a lesser extent, in fission yeast (Hennig et al., 2012; Kizer et al., 2005; Shim et al., 2012). Our results therefore confirm that Ser2-P, involved in Set2 recruitment, is required to enforce transcriptional fidelity in fission yeast.

In contrast, no global splicing defect was observed in the S2A mutant, in agreement with a recent publication (Inada et al., 2016). These results contrast with early studies demonstrating splicing defects of specific reporter introns in human and chicken cells containing S2A mutations (Gu et al., 2013; Hsin et al., 2014), but might be consistent with a more recent report suggesting that the spliceosome is associated primarily with Ser5-P and not Ser2-P (Nojima et al., 2015). Taken together, our results and these studies argue that Ser2-P is not globally required in coordinating splicing with transcription in fission yeast, an absence of requirement that might be conserved in higher eukaryotes as well, although the situation is more controversial.

As for the gene-specific requirements, we confirmed that only a fraction of the protein-coding genes (about 10%) requires Ser2-P for proper expression, including the *ste11* master regulator of sexual differentiation, as expected from our earlier studies (Coudreuse et al., 2010; Materne et al., 2015; Materne et al., 2016). Besides the protein-coding genes, an unexpected class of genes were found to be differentially expressed (upregulated) in the S2A mutant: the Pol III-transcribed tRNA genes.

ii. Mechanisms for tRNA accumulation in the S2A mutants.

To understand how the loss of Pol II Ser2-P could affect Pol III transcriptional output, we performed a series of experiments and analysis whose results are summarized hereafter:

- No gene coding for protein involved in Pol III transcription or its regulation was differentially expressed in the S2A mutant. Consistent with the idea that Pol II could directly affect tRNA expression, Pol II, Ser2-P and Ser5-P are present at tDNAs. In addition Pol II is active in anti-sense of the tDNAs, although the product of that transcription is very unstable.

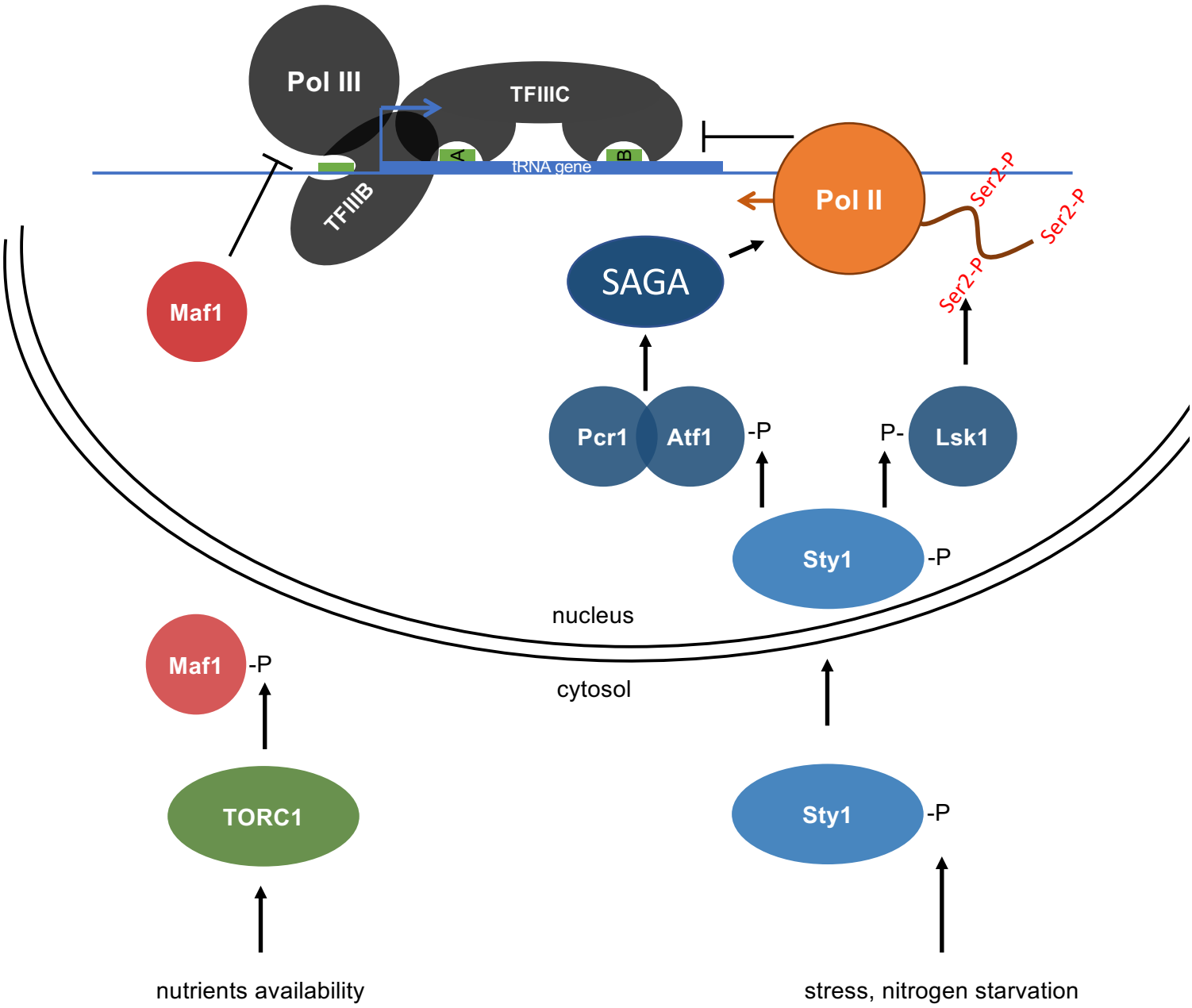


Figure 19: A new regulatory layer for Pol III transcriptional regulation ?

Maf1-dependent regulation of Pol III transcription: When nutrients are available, Maf1 is phosphorylated by TORC1 and other kinases. In consequence, it localizes in the cytoplasm, preventing its repressive activity. However, in response to unfavorable conditions, Maf1 is dephosphorylated and is allowed to enter the nucleus, where it represses Pol III transcription

Model for a Pol II-dependent regulation of Pol III transcription: Pol II is present at tRNA genes at a basal level. Upon stress, the Sty1 MAP kinase is phosphorylated and shuttle in the nucleus, where it phosphorylates Lsk1 and Atf1. Phosphorylated Lsk1 boosts Ser2-P phosphorylation, while phosphorylated Atf1 could favor SAGA recruitment at tRNA genes, further activating Pol II transcription. In consequence, Pol II possibly interferes more with Pol III transcription, and negatively regulate through a currently unknown mechanism.

- The chromatin structure at tDNAs is dependent on Ser2-P. Indeed, in the S2A mutant, a fragile MNase-sensitive particle becomes more resistant to MNase degradation. However, the nature of the MNase-sensitive is currently unclear.
- Despite tRNA accumulation in the S2A mutant, Pol III is actually less present at tRNA genes in that mutant. Therefore, tRNA accumulation could be due to the stabilization of the tRNA precursors. Accordingly, deletion of the exosome subunit *rrp6* causes massive tRNA accumulation. However, we could not highlight a connection between the exosome function and/or recruitment and Ser2-P.

All in all, the mechanistic connection between Ser2-P and tRNA accumulation remains elusive. In particular, the discrepancy between the increased tRNA level and decreased Pol III occupancy observed in the S2A mutant is difficult to interpret. One possible explanation could that Ser2-P affects tRNA expression on two level, affecting both Pol III occupancy and tRNA stability in opposite ways.

Nevertheless, the correlation between Pol II activity at tRNA genes and tRNA accumulation in the S2A mutant is suggestive of a direct effect. Indeed, it is hard to conceive that anti-sense Pol II transcription at tRNA genes would not, one way or another, interfere with sense Pol III transcription. As Pol II was also found in association with actively transcribed Pol III genes in mammals – reviewed in (White, 2011) – we can speculate that such Pol II-Pol III interferences might be widespread in eukaryotes.

iii. A new regulatory layer to control tRNA expression.

Strikingly, the S2A mutant appeared unable to properly shut down tRNA expression in conditions of nutritional stress, in a *maf1*-independent way. Suggesting that both Maf1 and Ser2-P might act in parallel to repress tRNA expression in stress conditions, our preliminary results showed that the double *Δmaf1 S2A* mutant have a somewhat aggravated growth defect phenotype, especially when recovering from long-term starvation.

In addition, in a screen without *a priori* for suppressors of a Pol III-transcriptional defect, we uncovered that disruptive mutations within the SAGA complex or the Pcr1 transcription factor lead to a clear suppression of the growth defect associated with the Pol III transcriptional defect. Strikingly, as Pcr1 and SAGA are associated with the regulation of Pol II transcription, this result further connects Pol II with Pol III transcription.

Intriguingly, all those newly identified player associated with the control of tRNA expression (or at least, suppressive of a Pol III growth defect) are part of the same regulatory pathway. Indeed, Pcr1 is a transcription factor that cooperates with its heterodimer partner Atf1 to recruit SAGA at gene promoters (Sansó et al., 2011). Atf1, but also the Ser2 kinase Lsk1, are phosphorylated by the Sty1 MAP kinase in response to stresses – such as nitrogen starvation (Salat-Canela et al., 2017; Sukegawa et al., 2011). The phosphorylation of Lsk1 leads to an increase in Ser2-P (Materne et al., 2015; Sukegawa et al., 2011), while Atf1 phosphorylation favors its transcription factor activity (Wilkinson et al., 1996) and SAGA mobilization (Sansó et al., 2011).

Together, even if we might face very indirect effects, our results suggest that an alternative pathway to the Maf1-dependent repression of tRNA expression exists, a pathway possibly dependent on Pol II transcription at tRNA genes and connected to environmental cues via the Sty1 MAP kinase (**figure 19**). However, more work is required to validate this model and to assess how, exactly, Pol II transcription and Ser2-P at tRNA genes interfere with tRNA expression.

PART 2: PHOSPHORYLATION OF SER2 ON THE RNA POLYMERASE II CTD IS REQUIRED FOR EFFICIENT TRANS-SPLICING OF POLYCISTRONIC TRANSCRIPTS IN *C. ELEGANS*.

Contributions:

RNA extractions, cdk-12as design and phenotyping, Figure 2:	Clement Cassart
ChIP extractions, immunofluorescence, Figures 1.B-C:	Fanélie Bauer
ChIP-seq and RNA-seq libraries generation and analysis:	Carlo Yague-Sanz

1. INTRODUCTION

The Ser2 phosphorylation (Ser2-P) is a widespread and abundant RNA polymerase II CTD modification that is conserved in all eukaryotes. Despite being observed on most Pol II genes, our work in the *yeast* model organism *S. pombe* shows that Ser2-P is dispensable for growth and for the proper expression of most genes (see the first part of this thesis; Coudreuse et al., 2010; Materne et al., 2015). As detailed in the general introduction, others studies in mammalian cell systems, fruit fly and budding yeast, also reached the conclusion that Ser2-P only required for the expression of subsets of genes (Cassart et al., 2010; Blazek et al., 2011; Ekumi et al., 2015; Pan et al., 2015; Li et al., 2016).

While working on single-celled organisms such as yeast is convenient for reasons mentioned in the general introduction, it cannot provide relevant insight on the complex regulation of the development of multicellular organisms. To study the potential implication of Ser2-P in developmental processes, we therefore choose a well-described multi-cellular organism: *Caenorhabditis elegans*.

A. *C. elegans* development.

In optimal conditions, *C. elegans* has a short life cycle (convenient for lab work) of about three days. The adults, which are in large majority hermaphrodites – 99.8% of the population are hermaphrodites, the remaining 0.2% being males (Klass et al., 1976) –, are able to self-fertilize their eggs, in which embryos develop during approximately 12h into L1 larvae. Once the L1 larva is formed, the egg hatches and the larva starts eating bacteria in its environment. It accumulates mass and passes through three additional larval stages (L2, L3 and L4) until it reaches adulthood about two days after hatching. An adult *C.*

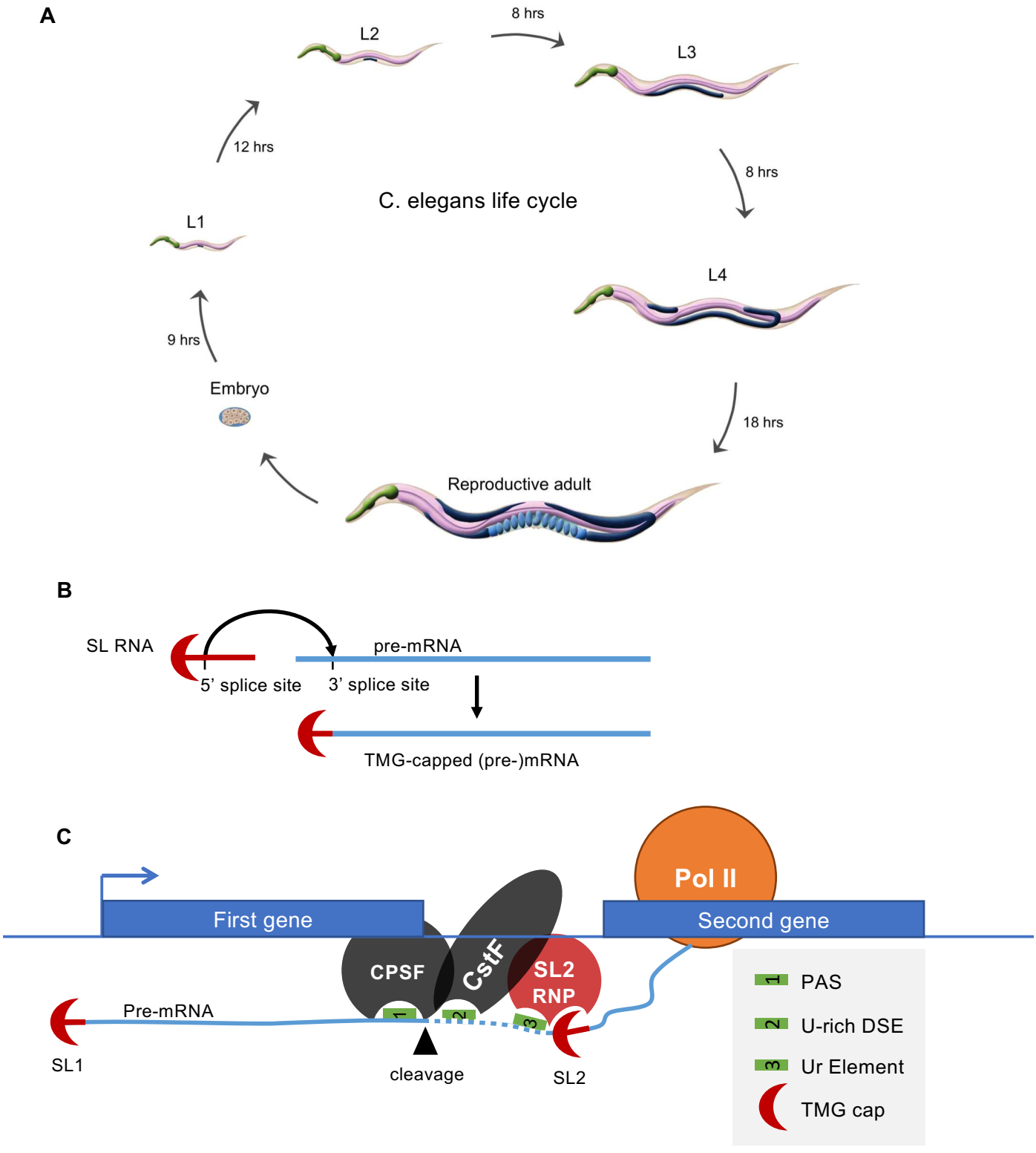


Figure 0: *C. elegans* life cycle, trans-splicing and operons

A. The C. elegans life cycle. Modified from (Wolkow et al., 2016). See main text for details.

B. The SL trans-splicing process. See main text for details

C. Coordination of the SL2 trans-splicing with the processing of polycistronic transcripts. See main text for details. CPSF: cleavage and polyadenylation specificity factor; PAS: polyadenylation signal; DSE: down-stream element. RNP: ribonucleoprotein; CstF: cleavage stimulatory factor; TMG: trimethylated guanosine.

elegans worm has a lifespan of about two weeks and is able to produce up to 300 eggs during that period of time (Byerly et al., 1976) (**figure 0.A**).

However, in an always changing environment, bacteria appropriate for C. elegans diet can be unavailable. If the L1 larvae hatch in such environment, it enters a developmental arrest state called "L1 arrest". During starvation, the L1-arrested larvae are able to survive for up to four weeks – nearly twice the normal lifespan for a fed individual –, slowing down their metabolism and building up their stress resistance (Baugh, 2013). Once food is available in the environment, L1-arrested larvae rapidly recover from starvation and resume their development. Remarkably, the marks of ageing accumulated by the starved L1 larvae are reversed upon the return to development induced by feeding. In consequence, the time spend during the L1-arrested stage does not affect adult lifespan (Roux et al., 2016).

B. Operons and trans-splicing in *C. elegans*.

i. Operons

A peculiarity of *C. elegans* genome – which was the first genome of a multicellular organism to be sequenced (Consortium, 1998) – is to contain operons. Operons are common in prokaryotes, but usually sparse in eukaryotes, except in some phyla including nematodes and trypanosomes. In the *C. elegans* nematode, there are 1371 annotated operons that contain clusters of two to eight genes whose transcription starts from a single promoter (Blumenthal et al., 2002). In contrast with bacterial operons, the genes within the same operon *in C. elegans* are not part of the same pathway and the polycistronic pre-mRNA is co-transcriptionally processed into monocistronic mature mRNAs. Therefore, the evolutive advantage of such operon structures can be difficult to understand. However, based on the observation that many genes organized in operons are over-expressed during the recovery phase post L1 starvation, it has been proposed that clustering these genes within operon can allow to more effectively transcribe them in recovery. Indeed, after starvation, the cell resources are sparse and the number Pol II transcriptional machinery available can be limiting. In the case of operons, only one Pol II transcriptional machinery is required to transcribe the whole operon, which is more efficient in such conditions with limited Pol II transcriptional machineries. In consequence of this optimization, the worms can proceed more rapidly through the transcriptional program of recovery and resume their development faster, which surely confers an evolutionary advantage (Zaslaver et al., 2011).

ii. Trans-splicing.

The co-transcriptional processing of the operon-born polycistronic pre-mRNA into monocistronic mature mRNA depends on the co-transcriptional 3' end processing (cleavage and polyadenylation) of the transcript originating from the upstream genes, coupled with a process called "spliced leader 2 trans-splicing" on the transcript originating from the downstream genes (Blumenthal et al., 2015). Spliced leader 2 (SL2) trans-splicing consists in the splicing of a capped small nuclear RNA called "spliced leader 2" onto the 5' end of a pre-mRNA molecule, substituting for canonical capping. This process, also catalyzed by the spliceosome (Hannon et al., 1991), differs from canonical

"cis-"splicing by the fact that the exons spliced together come from two distinct RNA molecules (**figure 0.B**).

While the SL2 trans-splicing is restricted to the transcripts originating from genes in position two and over in operons (about 10% of the annotated protein-coding genes), another type of SL trans-splicing can occur on a majority of the remaining pre-mRNA transcripts originating from promoter-proximal genes: the SL1 trans-splicing, a process that uses the SL1 RNA. Intriguingly, some transcripts whose gene is located within operons containing facultative internal promoters can be trans-spliced by either SL1 or SL2 depending on whether the transcription initiated at the internal promoter or at the operon promoter (Allen et al., 2011). The function of the SL1 trans-splicing remained elusive until recently, when it has been attributed to the enhancement of translational efficiency (Yang et al., 2017)

Specific signals converge to provide the specificity of the recruitment of the SL2 particle to the transcripts from the genes downstream in operons. The cleavage and polyadenylation specificity factor (CPSF) recognizes the polyadenylation signal (PAS) on the upstream pre-mRNA. Then, both a uridine-rich downstream element (U-rich DSE) and the CPSF, participate in the recruitment of the cleavage stimulatory factor (CstF) through its Cstf-50 subunit. Finally, the SL2 ribonucleoprotein (RNP), containing the spliceosome loaded with the SL2 RNA, is recruited by both the Ur Element and CstF and catalyzes the timely trans-splicing of the capped SL2 RNA to the second pre-mRNA (Graber et al., 2007) (**figure 0.C**).

C. CTD-Ser2 kinases in *C. elegans*.

In *C. elegans*, there are two CTD-Ser2 kinases: CDK-9 and CDK-12. CDK-9 was the first one identified and is essential for the embryonic development of the worm (Shim et al., 2002). Upon CDK-9 knock-down, global mRNA level is strongly reduced and Ser2-P becomes undetectable in all cells of the embryo, except in the two primordial germ cells[8] (Furuhashi et al., 2010). In contrast, CDK-12 knock-down leads to a strong decrease of Ser2-P in all cells (with residual phosphorylation still detected), but the embryos are viable and develop until the L1 larval stage (Bowman et al., 2013).

So, which kinase is the main CTD-Ser2 kinase in *C. elegans*? One proposed interpretation of the aforementioned results is that there is a tissue-specific regulation of the Ser2 phosphorylation: in somatic cells, some Ser2-P by CDK-9 would prime the CDK-12 activity, while in the germ line cells CDK-12 is active independently. In that model, while both kinases are required for bulk Ser2 phosphorylation, CDK-12 is directly responsible for most of it and can be considered as the main CTD-Ser2 kinase.

[8] Primordial germ cells are two cells (called Z2 and Z3) present in the L1 worm after hatching that will start proliferate mid-L1 and form the eggplant-shaped germline over the course of repeated meiosis (Hubbard et al., 2005).

D. Objectives

From these seminal studies, it appears that CDK-12-dependent Ser2-P is not required for the transcription and worm embryonic development. Instead, it is essential for larval development from stage L1 to L2 (Bowman et al., 2013). Interestingly, this L1 – L2 transition is tightly regulated in response to the presence of food (bacteria) in the worm environment: in starvation, the larvae remain in L1 stage for several weeks until they die. However, when food is available, they start recovering from starvation, rapidly changing their transcriptional program to develop into L2 larvae (Maxwell et al., 2012).

In this work, we studied how CDK-12 – and, by extension, Ser2-P – is connected to the worm development beyond the L1 stage. Using both RNA-seq and ChIP-seq, we found that only a subset of genes is sensitive to the loss (or at least, a strong decrease) of Ser2-P – as it is the case in yeast. Consistent with the L1-arrest phenotype of CDK-12 inhibition, the genes affected comprised genes annotated as being involved in development. Importantly, the majority of CDK-12-dependent genes are also organized in operon, revealing a role for CDK-12 in coupling transcription with the processing of polycistronic transcripts into discrete mRNAs. Finally, we discuss how the efficiency of this co-transcriptional processing, which appears CDK-12-dependent, might be tuned in order to tightly control gene expression within the operon.

2. Results & Discussion

A. An analog-sensitive allele of the Ser2 kinase Cdk-12.

In order to study the functions of the essential CDK-12 protein, we had to find a system to conveniently inactivate it, either at the gene level, the mRNA level or protein level. Several options were available, including:

- **Gene disruption:** the *tm3846* allele (from the National BioRessource Project) is a partial deletion in the *cdk-12* gene, which causes a frame-shift. Homozygous mutants for this allele are arrested in L1 and, therefore, cannot be maintained (Bowman et al., 2013).
- **Gene silencing**: RNAi against *cdk-12* also results in L1 arrest (Bowman et al., 2013) while allowing for more control as for when the inhibition takes place. However, it has some caveats: the inhibition is not always reproductible between experiments and can be experimentally challenging.
- **Kinase inhibition:** Because of those caveats, we wanted to create a *cdk-12* conditional mutant using the Shokat method (Alaimo et al., 2011). It consists in the introduction of a mutation in a conserved residue of the kinase catalytic pocket. In consequence, the pocket becomes slightly larger. While still being able to bind and catalyze ATP, the kinase with the enlarged pocket can also bind synthetic ATP-analogs too large for natural kinases. Those ATP-analogs cannot be hydrolyzed and therefore act as "suicide substrate", effectively inhibiting the mutated kinase in a very efficient, specific, reproducible, rapid, and reversible way (**figure 1.A**).

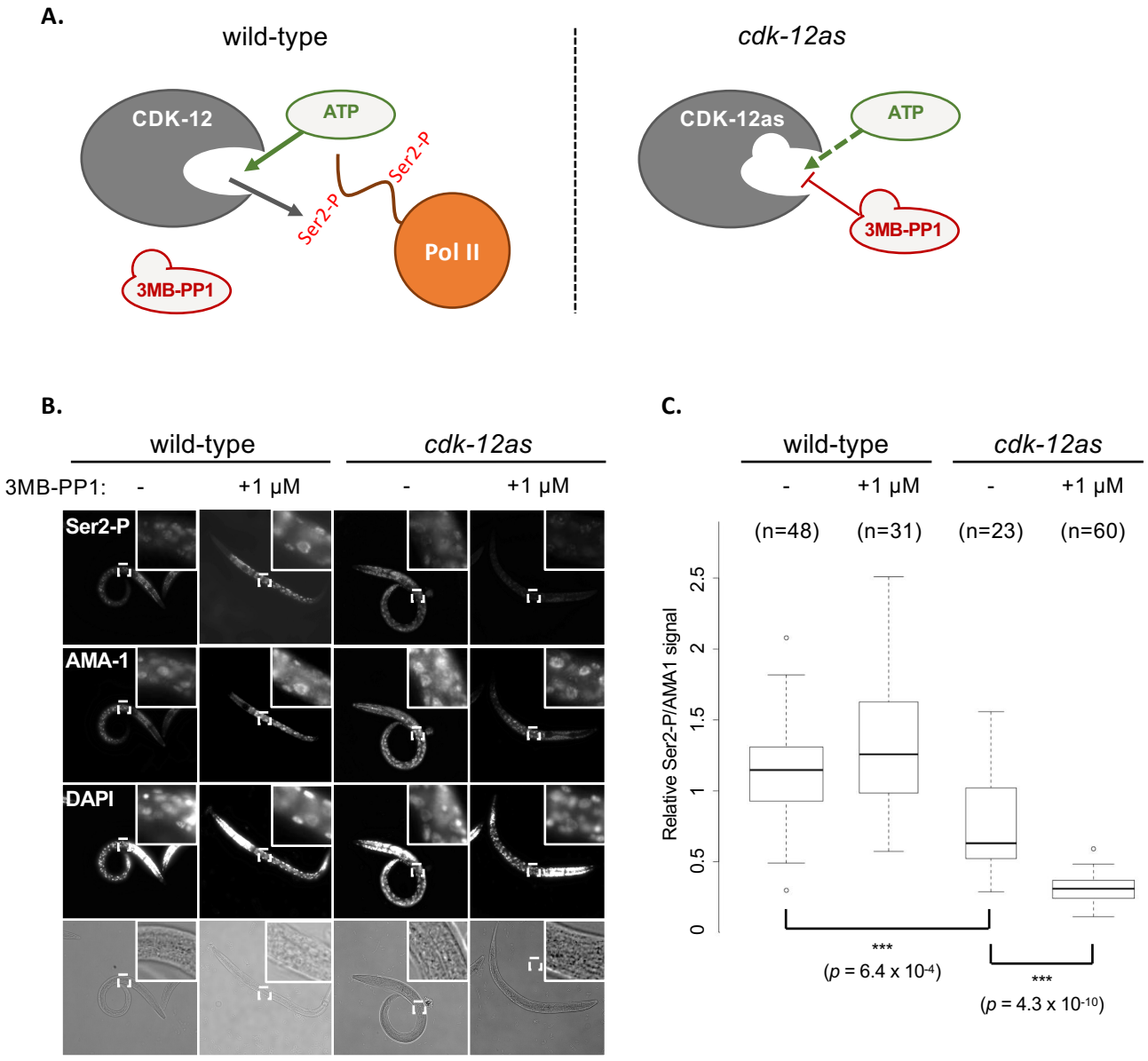


Figure 1: An analog-sensitive mutant of the Ser2 kinase CDK-12.

A. Cartoon representation of *C. elegans* CDK-12 and CDK-12as proteins. CDK-12 (wild-type) is a kinase of the Pol II CTD and uses ATP to phosphorylate Ser2. Bulky ATP analogs, such as 3MB-PP1, cannot be bound by the kinase (or any other natural kinase) because of steric hindrance. In contrast, CDK-12as carries a mutation that enlarges the catalytic pocket of the kinase. While it can still bind and catalyze ATP, it is also able to bind 3MB-PP1, which cannot be catalyzed and therefore blocks the kinase activity by acting as a "suicide substrate".

B. Immunofluorescence of wild-type and *cdk-12as* L1 worms treated with either DMSO or 1µM 3MB-PP1 and fed for 4h. DAPI (that colors nuclei) and antibodies against Ser2-P and AMA-1 (Pol II largest subunit that contains the CTD) were used. Inserts = 5x zoom on the gut region close to the pharynx.

B. Quantification of Ser2-P immunofluorescence signal relative to AMA1 in whole worms (n>22). P-values and significance of paired Wilcoxon-Mann-Whitney tests are indicated.

Given its advantages, this last method was chosen to study CDK-12 function. The strain mutated for the kinase was obtained and called *cdk-12as* (*"as"* standing for *"analog sensitive"*) (see Clement Cassart's thesis). The inhibition was performed using the ATP analog 3MB-PP1, known to be the most effective inhibitor for *as* kinases (Zhang et al., 2013). After CDK-12 inhibition, Ser2-P levels dropped dramatically within the cell nucleus as seen by immuno-staining of L1 worms (**figure 1.B-C**). This validates the efficiency of our conditional system for CDK-12 inhibition and confirms that CDK-12 is a genuine CTD-Ser2 kinase. However, it should be noted that even without inhibitor, the level of Ser2-P in the *cdk-12as* strain is lower than in the wild-type. This suggests that *cdk-12as* is a hypomorphic allele, probably because the introduced mutation causes a partial loss of function of CDK-12, leading to an intermediate level of Ser2-P. As expected, the addition of inhibitor does not decrease Ser2-P levels in the *wild-type* control worms – Ser2-P level actually increases, possibly because of technical variations as the immunofluorescence on L1 worms is experimentally challenging –, confirming that the activity of CDK-12 is not inhibited by 3MB-PP1 when the kinase is not sensitized by the *as* mutation.

B. Cdk-12 is required for post-L1 development.

Despite the *cdk-12as* allele being hypomorphic, the worms are able to develop past the L1 stage like wild-type worms. However, upon CDK-12 inhibition starting at the L4 stage with the specific 3-MB-PP1 inhibitor, only L1 larvae are observed on the plate from the progeny, indicating that they cannot develop further into the L2 stage (**figure 2**). This phenotype, mimicking the starvation-dependent L1-arrest, is consistent with the results from a previous study, obtained with a homozygous mutant for a *cdk-12* null allele or with RNAi targeted against CDK-12 (Bowman et al., 2013).

Intriguingly, CDK-12 appears to be required for post-L1 development, but not for embryonic development, while both stages require active transcription (Gerstein et al., 2010) that is accompanied by CTD-Ser2 phosphorylation by CDK-12. To explain this difference in CDK-12-dependence, we hypothesized that Ser2-P could be essential for the transcription of a subset of genes (such as genes involved in the L1-L2 transition) but dispensable for the transcription of others.

C. Cdk-12 inhibition impairs the expression of a subset of development genes.

To test this hypothesis, we decided to analyze the transcriptome of L1 *cdk-12as* worms during the early stages of this L1-L2 transition. We studied the wild-type and *cdk-12as* strains in four conditions that we named "S0", "S", "R" or "R+I". In the "S0" (the "S" standing for *starved*), the freshly hatched worms were starved for 12h, being effectively synchronized in the L1 stage. From there, the population was divided in three conditions:

- In the "S" (*starved*) condition, the L1-arrested larvae from the S0 condition were **starved** for four additional hours.
- In the "R" (*recovery*) condition, the L1-arrested larvae from the S0 condition were **fed** during four additional hours.

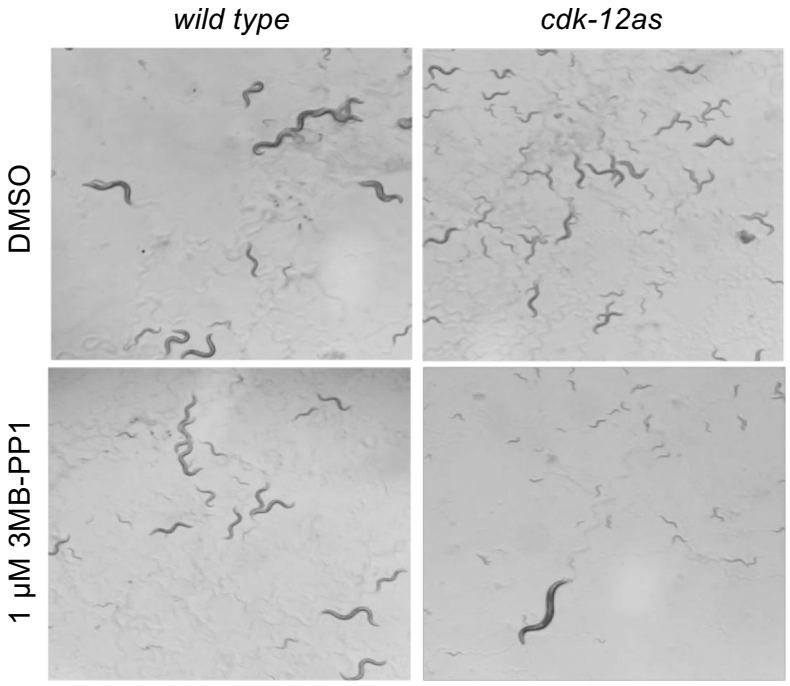


Figure 2: CDK-12 activity is dispensable for embryonic development, but required for post-L1 development.

L4 worms were grown in liquid cultures in presence of 1µM 3MB-PP1 or DMSO. Three days later, we observed progeny as old as L4 in the wild type worms and *cdk-12as* worms without inhibitor while only the parental worm with its L1-arrested progeny were observed for the *cdk-12as* strain inhibited with 1µM 3MB-PP1.

- In the "R + i" (*recovery + inhibitor*) condition, the L1-arrested larvae from the S0 condition were **fed** as in the "R" condition, but in presence of 1 µM of 3-MB-PP1 **inhibitor**.

Finally, as the 3MB-PP1 inhibitor is resuspended in dimethyl sulfoxide (DMSO), and that DMSO can affect gene expression (Wang et al., 2010), equal volume of DMSO was added in all conditions without inhibitor, except from the additional L1-starved control ("*S0*") taken before DMSO addition.

For those "S0", "S", "R" and "R + I" conditions, summarized in **table1**, we sequenced two replicates (in two batches) of ribo-depleted RNA extracts using the Illumina TruSeq protocol coupled with HiSEQ paired-end 50 nt sequencing to an depth of coverage between 18 and 40x. Because fed worms (*"R"* and *"R+ i"* conditions) tend to have bacteria stuck to their cuticle and within their pharynx and gut, reads were first mapped to the *E. coli* OP50 (the bacterial strain used to feed the worms) genome and excluded from further analysis. As expected, only the conditions with fed worms had significant levels of *E. coli* contamination while the residual contamination observed in the other conditions could be attributed to aspecific mapping of the reads on sequences homologous between *C. elegans* and *E. coli*. The high variability observed in the percentage of *E. coli* reads recovered between replicates 1 and 2 for the conditions with fed conditions could be attributed to variation in the efficiency of the washes performed on the full worms before RNA extraction.

Table1: Conditions studied by RNA-seq.

condition	strain	food	inhibitor	DMSO	batch	*E. coli* (‰)	depth[a]
S0	wild type	starved	Ø	Ø	1	0.2	28.5
S0	wild type	starved	Ø	Ø	2	0.1	19.3
S	wild type	starved	Ø	DMSO	1	0.3	37.2
S	wild type	starved	Ø	DMSO	2	0.1	22.7
R	wild type	fed	Ø	DMSO	1	7.0	24.8
R	wild type	fed	Ø	DMSO	2	0.7	23.5
R + i	wild type	fed	3MB-PP1	DMSO	1	16.0	38.2
R + i	wild type	fed	3MB-PP1	DMSO	2	0.9	25.9
S0	*cdk-12as*	starved	Ø	Ø	1	0.1	20.6
S0	*cdk-12as*	starved	Ø	Ø	2	0.2	37.1
S	*cdk-12as*	starved	Ø	DMSO	1	0.2	24.3
S	*cdk-12as*	starved	Ø	DMSO	2	0.1	29.1
R	*cdk-12as*	fed	Ø	DMSO	1	13.0	18
R	*cdk-12as*	fed	Ø	DMSO	2	2.5	39.5
R + i	*cdk-12as*	fed	3MB-PP1	DMSO	1	17.5	21.7
R + i	*cdk-12as*	fed	3MB-PP1	DMSO	2	6.2	39.3

a. The depth is calculated based on the number of uniquely mapped reads to the *C. elegans* genome divided by the length of the genome size (100 x 10^6 nt) multiplied by the read length (50 nt).

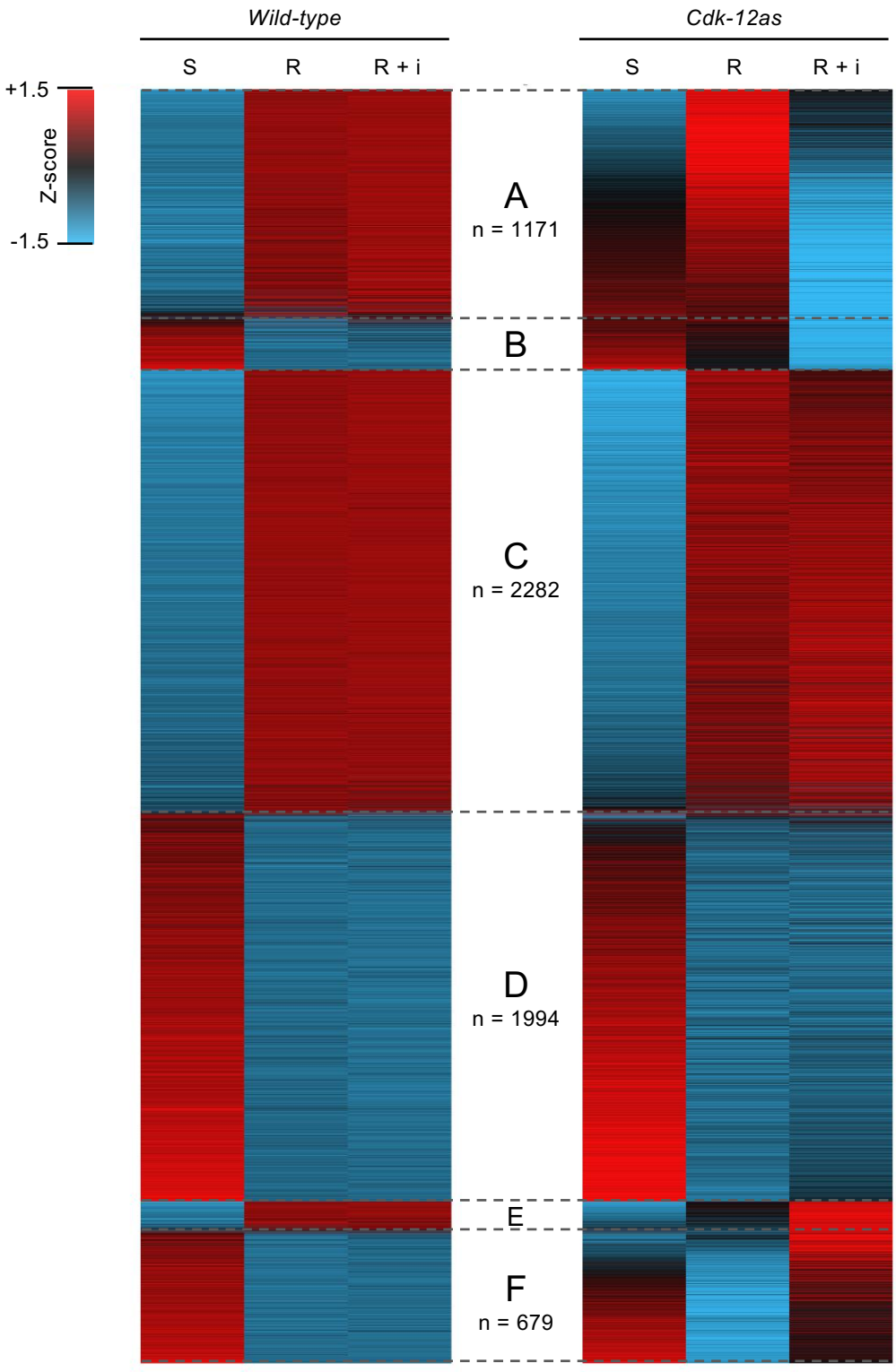


Figure 3: A subset of genes cannot be induced in recovery when CDK-12 is inhibited.

Heatmap of the Z-score of the rlog-transformed expression score averaged over the two replicates for the 6550 genes differentially expressed by the presence of food or inhibitor in either strain. Six genes clusters (A-F) are defined according to the food and inhibitor effect. "S" = starvation (no food); "R" = recovery (food was added during 4h); "R + i" = recovery + 1 µM 3MB-PP1 inhibitor. The number of genes in each cluster is indicated on the figure except for clusters B (271 genes) and E (153 genes).

The remaining reads mapped on the *C. elegans* genome were quantified at the gene level (i.e, on all annotated exons for a gene, without regard for alternative splicing) and multifactorial differential gene expression analysis was performed in DESeq2 using the following generalized linear model:

EXPRESSION ~ STRAIN + INHIBITOR + FOOD + STRAIN:INHIBITOR + STRAIN:FOOD + DMSO + BATCH

The individual effects of the strain (*cdk-12as* VS wild-type), inhibitor (3MB-PP1 vs Ø) and food (fed vs starved) were evaluated. In addition, since the inhibitor is expected to differentially affect gene expression in both strains and because the *cdk-12as* strain is not perfectly wild-type, we also included the interaction between the 'strain' and 'inhibitor' factors and between the 'strain' and 'food' factors. The 'DMSO' and 'batch' factors were not directly analyzed, but were integrated into the model as covariates in order to control for technical variation (batch effect) and the effect of DMSO[9] on gene expression.

To provide a general overview of the impact of CDK-12 inhibition in the context of the L1-L2 transition during recovery from starvation, we selected the 6550 genes differentially expressed by the 'food' or 'inhibitor' effect in either strains and draw a heatmap of gene expression (**figure 3**). Overall, the transition from starvation (*"S"*) into recovery (*"R"*) has a dramatic impact on the expression profiles, regardless of the genetic background. In contrast, the addition of the inhibitor (*"R + i"*) has, as expected, very little impact on the wild-type, but affects gene expression in the *cdk-12as* strain as detailed below:

- Most genes (clusters C and D) are unaffected by CDK-12 inhibition and behave similarly than in the "R" condition. This is somewhat surprising given that the phenotype of inhibited worms matches that of L1-starved worms. It indicates that the transcriptional program of the recovery from starvation leading into post-L1 development is globally activated even if it cannot complete properly.
- However, a relatively big subset of genes (cluster A), whose expression is normally induced in recovery, are under-expressed when CDK-12 is inhibited. Some of them remain at the same expression level than in starvation, others reach intermediate levels between starvation and recovery, but most are actually more under-expressed than in starvation.
- Conversely, cluster F contains genes, whose expression is normally repressed in recovery, are over-expressed upon CDK-12 inhibition.

[9] In retrospect, there is a small flaw in the rather well-controlled design of our experiment: the DMSO effect is confounded with a "*+4h*" effect because the worms in the *"S0"* condition were sacrificed 4h before the others. In consequence, gene expression changes due to the DMSO *(+4h)* effect are difficult to interpret, which prompted us to not display the "S0" condition in our results. However, it remains perfectly safe to use the DMSO factor as a covariate; and including the "*S0*" condition in our analysis has the advantage to improve the gene dispersion estimate used within DESeq2, increasing the power of our experiment.

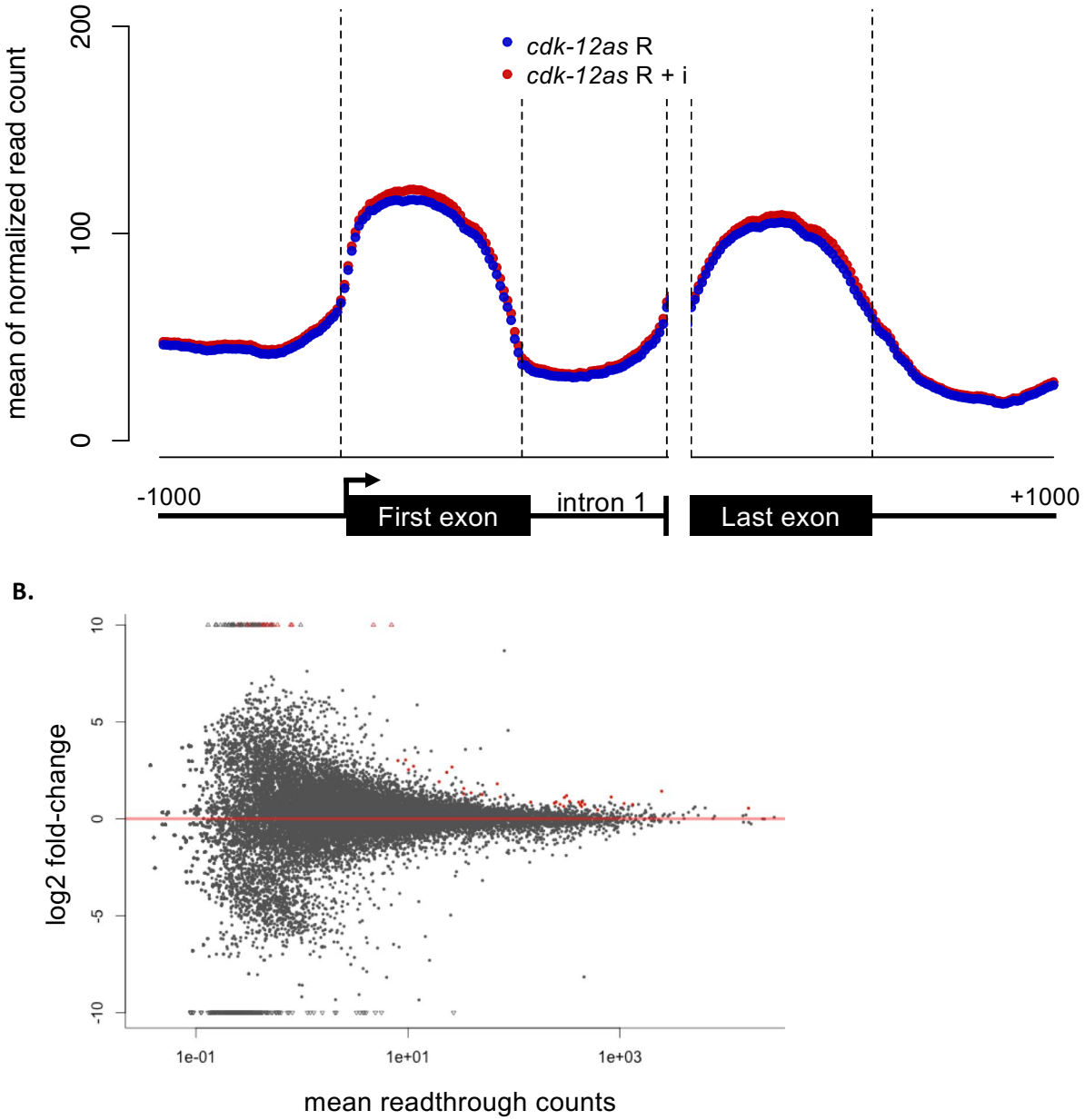


Figure 4: There is no global splicing or 3' end formation defect upon CDK-12 inhibition

A. Meta-gene of the mean of the normalized read count in the indicated conditions.

B. Differential read-through analysis upon CDK-12 inhibition. The number of reads mapped in a 500 bp window downstream the gene transcription end site (TES) in the *cdk-12as* R and R + i conditions were quantified and analyzed in DESeq2 to identify the genes with significantly more read-through upon CDK-12 inhibition (red dots). The genes for which there was another annotated gene on the same strand within a region of 1 Kb downstream of their TES were filtered out, along with the genes that were differentially expressed upon CDK-12 inhibition.

- Finally, the smaller clusters B and E contains genes that are respectively excessively under-expressed or excessively over-expressed in recovery when CDK-12 is inhibited.

Altogether, this analysis shows that only a subset of genes involved in recovery is misregulated upon CDK-12 inhibition, and that this misregulation is sufficient to phenotypically mimic the L1 arrest observed in starvation. Accordingly, 216 of the genes in clusters A are annotated by the gene ontology consortium (Ashburner et al., 2000) as being involved in nematode larval development. This does not constitute a specific enrichment – 582, 99, 17, 22 and 47 genes involved in nematode larval development are found respectively in clusters C, D, B, E and F – but could explain the observed phenotype. However, more detailed analyses were required to understand why some genes are affected by CDK-12 inhibition while others are not.

D. CDK-12 inhibition does not globally impact the pre-mRNA splicing or 3' end formation.

Only as relatively small subset of genes (1211 genes, including 1191 protein coding genes) were detected as differentially expressed (DE) upon CDK-12 inhibition (i. e., DE by the inhibitor effect in the *cdk-12as* strain). This is somewhat surprising given that Ser2-P deposited by CDK-12 is an ubiquitous mark, present on all expressed Pol II genes (Mayer et al., 2010), and has been implicated in essential co-transcriptional processes such as the splicing (Gu et al., 2013) and the cleavage and polyadenylation (Lunde et al., 2010) of the pre-mRNAs. However, despite this connection, those processes appear unaffected in mutants lacking Ser2-P in fission yeast as shown in the first part of this thesis and as recently published by another group (Inada et al., 2016).

Confirming that the impact of the CDK-12 inhibition in *C. elegans* is limited to the identified genes, we did not observe global splicing defect (intron retention) nor cleavage defect or readthrough (reads mapping beyond the gene boundaries) in the *cdk-12as* strain in the R + i condition (**figure 4.A**). Moreover, quantification of the "readthrough" reads in the 500 nt downstream the annotated transcription end sites revealed that only 53 genes (including 28 protein coding genes) have significantly more readthrough when CDK-12 is inhibited (**figure 4.B**). In addition, manual inspection of those genes in an online genome browser loaded with our RNA-seq tracks (https://tinyurl.com/ycvchkb4) showed rather anecdotical differences.

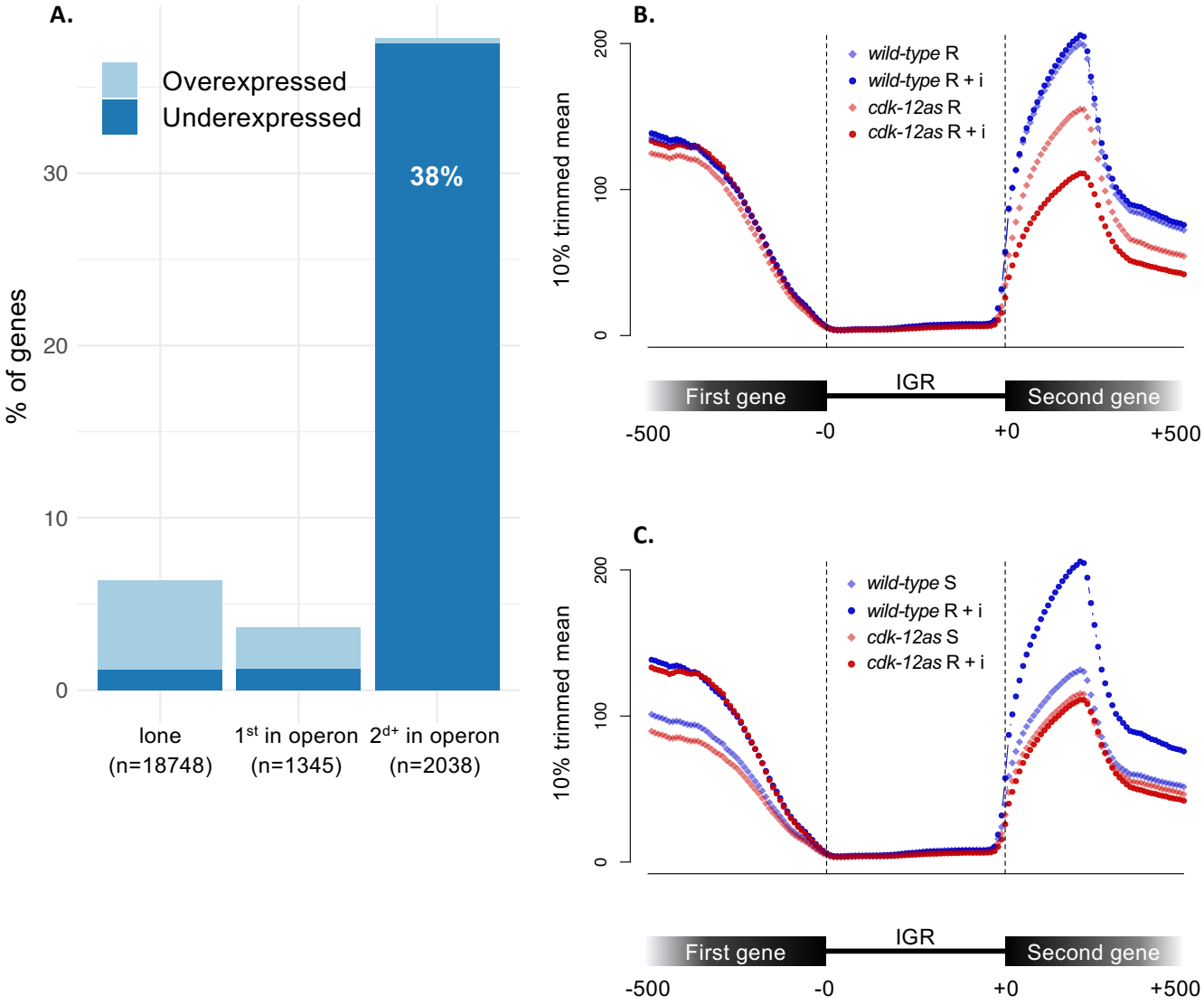


Figure 5: Genes under-expressed upon CDK-12 inhibition are located in position 2 and over in operons

A. Percentage of genes mis-regulated upon CDK-12 inhibition, either overexpressed (fold change < 0.66, FDR < 0.01) or underexpressed (fold change > 1.5, FDR < 0.01) stratified according to the gene localization in regards to operons. Lone genes: genes outside operons. 2^{d+} in operon: genes in position 2 and over in operons.

B - C. Meta-gene of the trimmed mean of normalized read counts centered on the scaled intergenic regions (IGR) between the first and second gene in operons in the indicated conditions. Only the 993 IGR whose length is smaller or equal to 800 were included in this analysis.

E. Genes in operons are specifically affected by CDK-12 inhibition.

Surprisingly, the genes affected by CDK-12 inhibition – in particular the genes down-regulated – are not randomly distributed within *C. elegans* genome: they are highly enriched in operons (clusters of genes transcribed from a single promoter), and, more specifically, in genes downstream of the first gene in the operons (Fisher's exact test pvalue < 2.2×10^{-16}) (**figure 5.A**). Metagene analysis confirmed that the first genes in operons are globally unaffected in the *cdk-12as* R + i condition while there is a global decrease of expression of almost two-fold for the genes downstream (**figure 5.B**). Interestingly, even without inhibition (*"R"* condition), those genes are expressed to an intermediate level in the hypomorphic allele *cdk-12as*, indicating that their expression is affected in a dose-dependent manner by CDK-12 activity and the level of Ser2-P.

Replacing these results in their biological context, we can see that the genes in operon are globally up-regulated in recovery from starvation in *wild-type strains* (**figure 5.C**). This is consistent with the findings of a published analysis of the transcriptional reprogramming occurring during the L1-L2 transition (Maxwell et al., 2012). In contrast, in the *cdk-12as* R + i condition, the expression in operon is properly induced for the first genes, but remains at the starvation level for the seconds genes

F. Cdk-12 is required for efficient SL2 trans-splicing of genes in operons.

By definition[10], the peculiarity of the genes in position two and over in operons is to be co-transcriptionally trans-spliced to the SL2 spliced-leader (SL) RNA (Allen et al., 2011). Therefore, we hypothesized that CDK-12 is directly involved in the SL2 trans-splicing. This hypothesis is supported by a recent study that found that Ser2-P peaks in the inter-genic regions between genes in operons (Garrido-Lecca et al., 2016).

To test the implication of CDK-12 in the SL2 trans-splicing, we quantified the SL-trans-splicing events in our RNA-seq dataset using a bioinformatics pipeline recently developed for this purpose: SL-quant (Yague-Sanz and Hermand, 2018). Briefly, the SL sequences are identified with high specificity and are trimmed from the input reads, which are then remapped on the reference genome and quantified at the gene level. The paper containing the details and validation of the method is attached in the **Appendix 2** (*Yague-Sanz and Hermand, 2018 – SL-quant: a fast and flexible pipeline to quantify spliced leader trans-splicing events from RNA-seq data"*).

Using this quantification, we calculated the proportion of SL2 trans-splicing events over the total number of trans-splicing events for each gene across all conditions. Stratification of the genes according to this *SL2/(SL1+SL2)* proportion revealed that the genes down-regulated upon CDK-12 inhibition are in majority mostly trans-spliced with

[10] Actually, annotation of genes in operons is directly based on their level of SL2 trans-splicing (Allen et al., 2011).

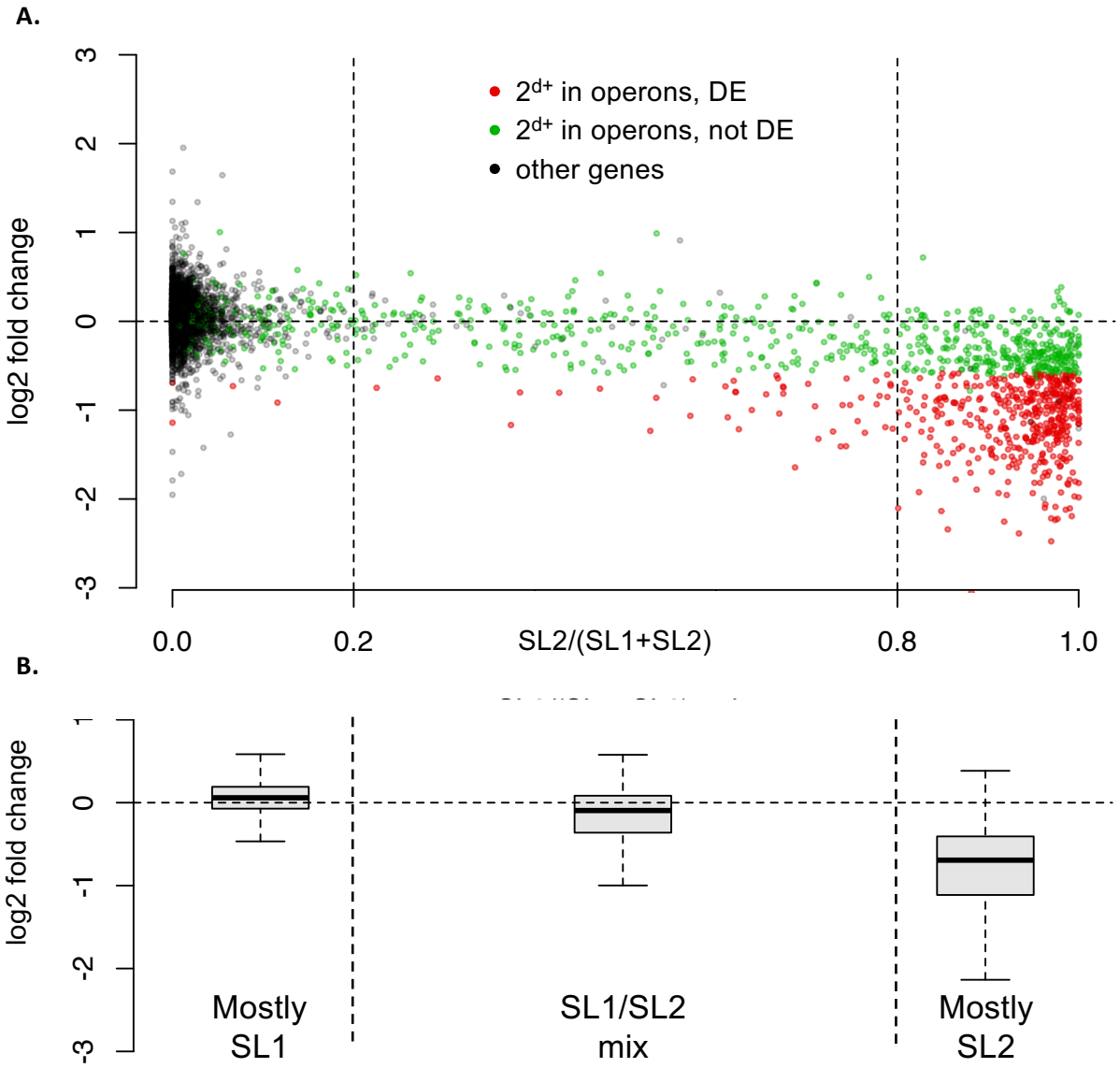


Figure 6: Genes in operon underexpressed upon CDK-12 inhibition are mostly trans-spliced with SL2.

A. Relationship between the gene fold change upon CDK-12 inhibition with proportion of SL2 trans-splicing events on the total number of trans-splicing events quantified with our SL-quant pipeline (see Appendix 1). Genes in position two and over (2^{d+}) in operons differentially expressed (DE) or not are respectively indicated in red and green. Genes with less than 64 total trans-splicing events detected were not included in the analysis.

B. Boxplot of the log2 fold change due to CDK-12 inhibition for genes classified according to their position in operon and their SL2/(SL1+SL2) ratio. Genes with less than 64 total trans-splicing events detected were not included in the analysis. *Mostly SL1:* genes with more than 80% SL1 trans-splicing (n= 3646). *Mostly SL2:* genes in operons with more than 80% SL2 trans-splicing (n= 728). *SL1/SL2 mix:* genes in operons in operons with between 20% and 80% SL2 trans-splicing (n= 294).

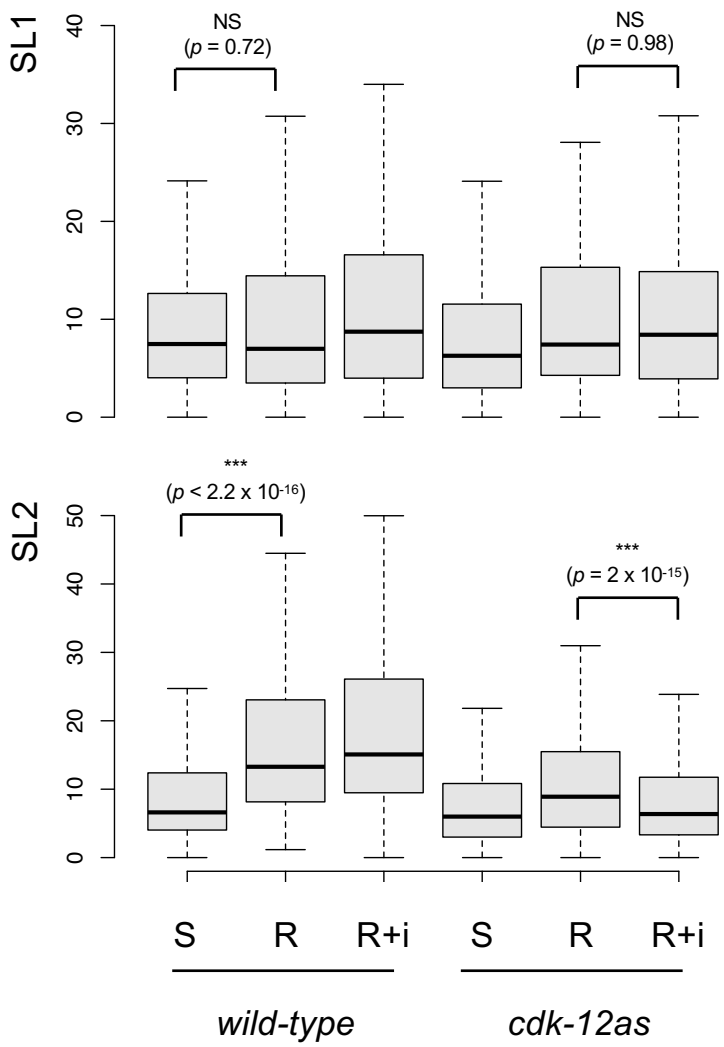


Figure 7: CDK-12 is specifically required for the SL2 trans-splicing of genes in operons.

Boxplot of the number of SL1 (top) and SL2 (bottom) trans-splicing events per gene in operons trans-spliced by both SL1 and SL2 (genes with more than 64 total SL trans-splicing events and between 20% and 80% of SL2) in the indicated conditions. (n=294). P-values and significance of paired Wilcoxon-Mann-Whitney tests are indicated.

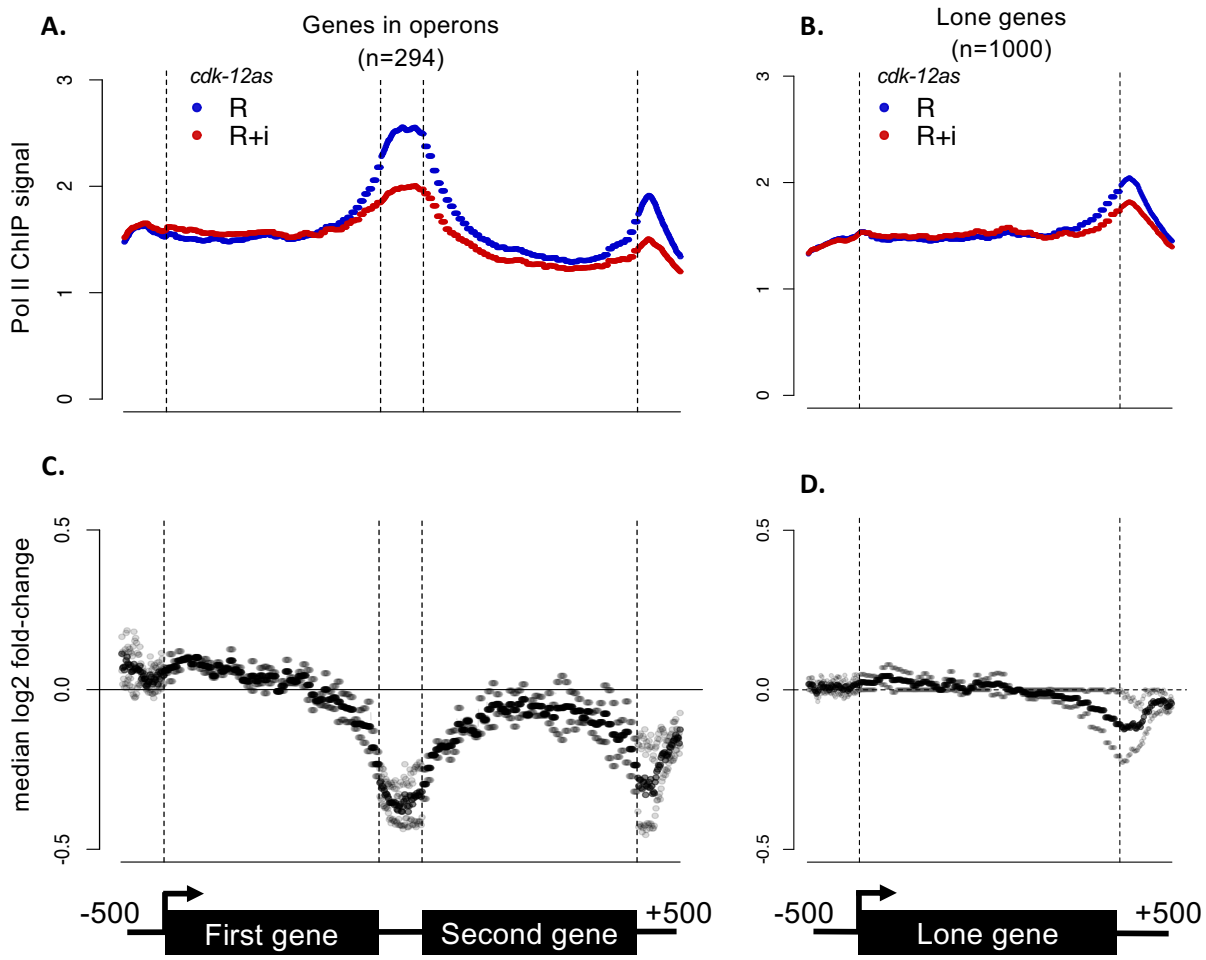


Figure 8: Pol II level on second genes in operons is decreased upon CDK-12 inhibition

A. Meta-operon profiles of BEADS-normalized Pol II occupancy data averaged over two replicates. Only the regions with one value above a threshold (5.058, set in such a way that only 1000 lone genes pass the threshold) were considered. Operons with an intergenic regions between the first and second genes bigger than 800 nt were also excluded from the analysis.

B. Meta-gene profiles of BEADS-normalized Pol II occupancy data averaged over two replicates. Only the regions with one value above a threshold (5.058, set in such a way that only 1000 lone genes pass the threshold) were considered.

C and D. Meta-gene profiles of the median log2 fold change between Pol II ChIP signal in the R and R + i conditions for the regions described in (A) and (B) respectively. Each replicate was processed separately (grey dots) and the average between both replicate for each data point is indicated with black doths.

SL2 (**figure 6.A-B**). In contrast, the effect of the inhibition is less pronounced for the genes in positions two and over in operons that can be trans-spliced by either SL2 (when the transcription starts from the promoter of the operon) or SL1 (when the transcription starts from an auxiliary internal promoter).

These results indicate that the gene-specific sensitivity to CDK-12 inhibition is not dictated by the gene localization within operons, but rather by the dependency to the SL2 trans-splicing for the co-transcriptional processing of the polycistronic pre-mRNAs.

To assess how CDK-12 inhibition affects the SL2 trans-splicing, we also compared the SL quantification between our conditions. In the *cdk-12as* R + i condition, we observed less SL2 trans-splicing than in the *cdk-12as* R condition (see for instance figure 11.A). However, this result is difficult to interpret because the decrease in SL2 trans-splicing could be the consequence, rather than the cause, of the down-regulation of the genes mostly trans-spliced with SL2. To disentangle those two effects, we took advantage of the genes trans-spliced with both SL2 and SL1. For those genes, the level of SL1 trans-splicing remains stable upon CDK-12 inhibition while the level of SL2 trans-splicing significantly decreases (**figure 7**), demonstrating that CDK-12 activity is required for efficient SL2 trans-splicing of the genes in operons.

G. Deficient SL2 trans-splicing leads to premature termination of Pol II transcription.

Next, we wondered about the fate of Pol II when CDK-12 is inhibited. We considered two main possibilities:

1. Pol II transcribes until the end of the operon, but the trans-splicing of the second (and over) genes in operons is impaired. Therefore, pre-mRNA lacking the capped SL2 snRNA are degraded.
2. The SL2 trans-splicing issue occurring between the first gene and the second (and over) genes in operon causes the polymerase to terminate prematurely.

To explore those possibilities, we made a Pol II ChIP-seq experiment for the conditions R and R + i in the *cdk-12as* strain. Overall, the efficiency of the immunoprecipitation was poor, perhaps because of technical difficulties of the chromatin extraction on starved (and then in recovery for 4h) L1 worms. Therefore, in order to obtain robust signal, we selected the genes with a BEADS-normalized signal (see material and methods) higher than a threshold defined in such a way that 1000 lone genes pass that threshold. Analysis of the meta-genes for those 1000 genes and for the genes in operons passing the same threshold reveals no difference upon CDK-12 inhibition on the body of the first genes in operon and on the body of lone genes (**figure 8.A**). In contrast, signal on the body of the second genes is slightly lower. More strikingly, Pol II signal in the intergenic region – that is more enriched probably because of Pol II slow down – between the first and the second genes and at the end of the second genes decreases dramatically.

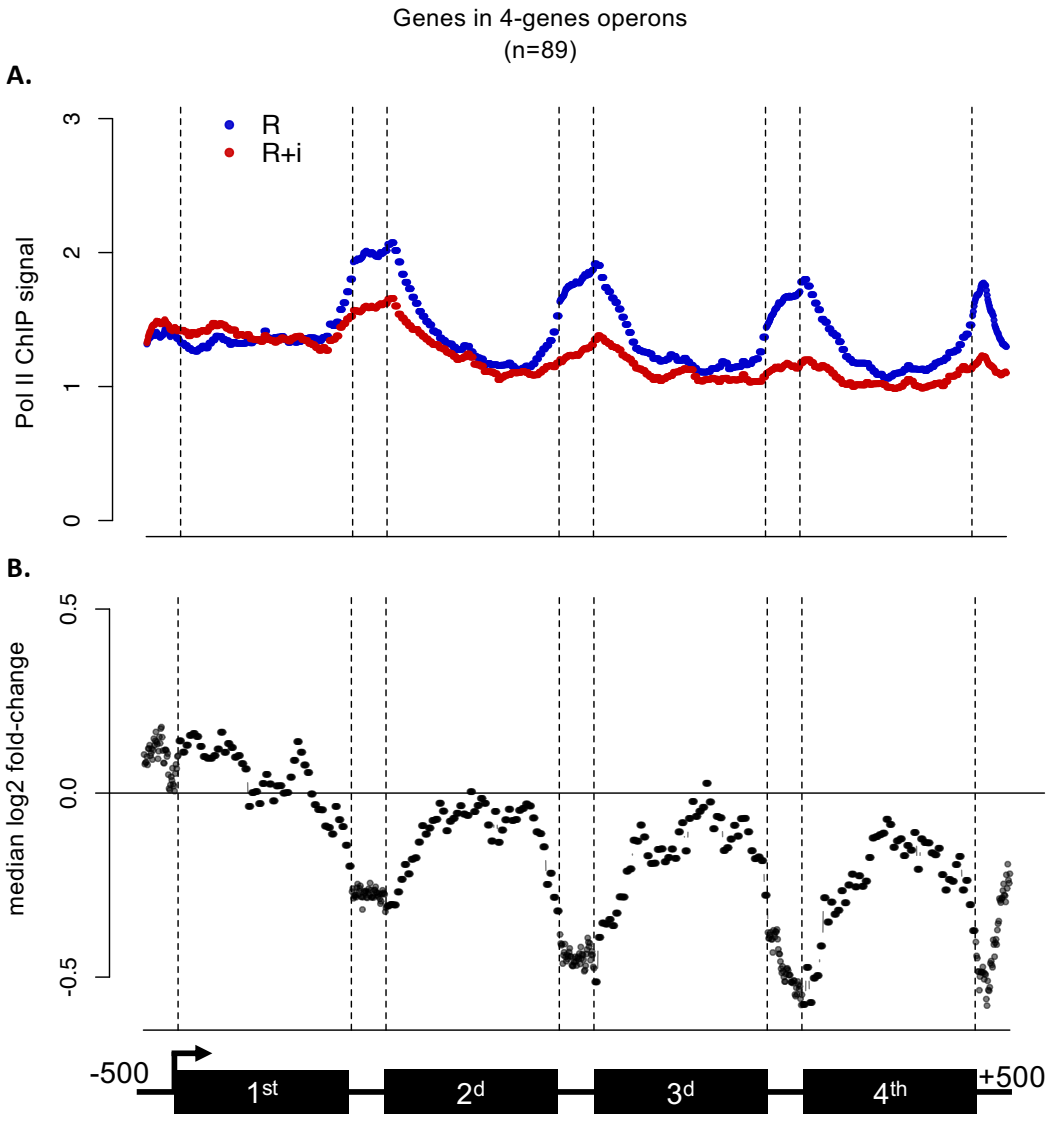


Figure 9: Pol II level decreases over the course of operon transcription when CDK-12 is inhibited.

A. Meta-operon profiles of BEADS-normalized Pol II occupancy data averaged over two replicates. Only the operons containing at least 4 genes were considered. Operons with any intergenic regions bigger than 800 nt were also excluded from the analysis.

B. Meta-gene profiles of the median log2 fold change between Pol II ChIP signal in the R and R + i conditions for the regions described in (A).

Overall, this analysis supports our second hypothesis – namely, that the trans-splicing defect causes the polymerase to terminate prematurely. To further challenge this hypothesis, we reasoned that if the SL2 trans-splicing failure truly causes the polymerase to terminate early, then for operons containing more than two genes, the effect would be amplified at each subsequent intergenic region. Confirming our hypothesis, meta-operon analysis of operons containing at least four genes revealed exactly the expected trend: the difference between the inhibited and normal condition grows larger over the course of the operon (**figure 9.A**).

Because meta-genes calculate average values between all genes, they are sensitive to extremes. In order to confirm our results and assess whether the individual operons behave similarly, we also plotted the median of the log2 fold change value between the two conditions at each position (**figure 8.B, 9.B**). At the end of the 4th genes in operons, the median log2 fold change reaches a minimum of about -0.6 (a difference of about 33%). Although robust, this difference is smaller than the difference of about two-fold (50%) we observed at the RNA level. It could be due to the relatively high level of noise in our ChIP-seq data, or this could mean that the two proposed hypotheses can coexist: Pol II sometimes terminates after a SL2-trans-splicing failure, but sometimes, it can go on.

Collectively, our results point toward the following model (**figure 10**): During the transcription of the operons, the cleavage and polyadenylation specificity factor (CPSF) recognizes the polyadenylation signal (PAS) on the first pre-mRNA. The cleavage of the pre-mRNA by CPSF creates an entry point for exonucleases. Therefore, to avoid degradation, the 5'end of the second mRNA needs to be rapidly capped. When CDK-12 is active, CTD-Ser2 is phosphorylated and participates, along with a uridine-rich downstream element (U-rich DSE) and the CPSF, in the recruitment of the cleavage stimulatory factor (CstF) through its Cstf-50 subunit. Finally, the SL2 ribonucleoprotein (RNP) is recruited by both the Ur Element and CstF and will catalyze the timely trans-splicing of the capped SL2 RNA to the second pre-mRNA.

However, when CDK-12 is inactivated, the recruitment of CstF and the SL2 RNP is less efficient due to the lack of Ser2-P. As a consequence, the SL2 trans-splicing does not occur (or is untimely) and the free 5' end of the second mRNA is be rapidly degraded by an exonuclease such as Xrn-2. When the exonuclease catches up with the polymerase, it leads to the destabilization of the template–transcript–Pol II ternary complex and ultimately causes transcriptional termination.

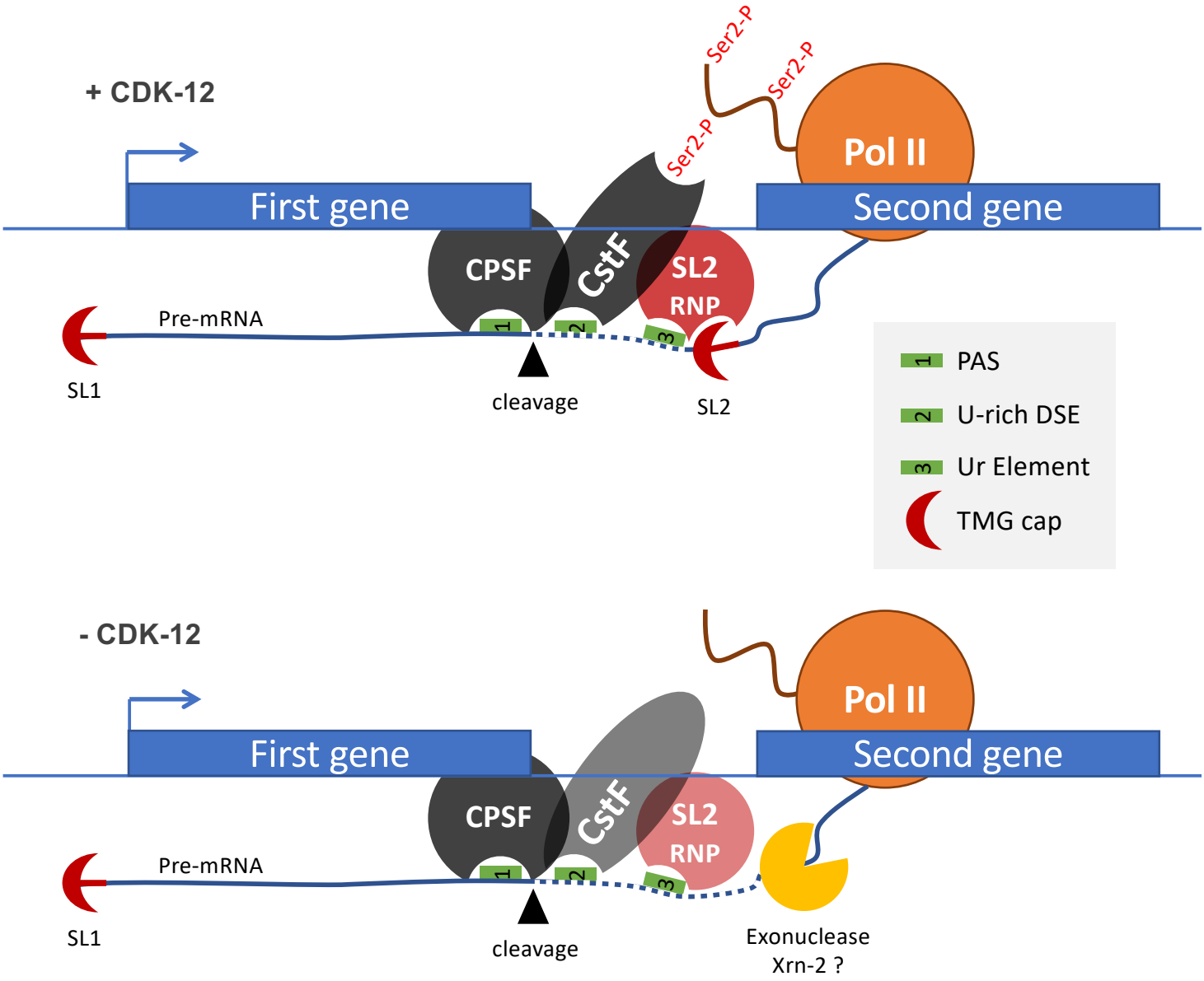


Figure 10: Model for the CDK-12 requirement in the SL2 trans-splicing process.

See main text for details. CPSF: cleavage and polyadenylation specificity factor; PAS: polyadenylation signal; DSE: down-stream element. RNP: ribonucleoprotein; CstF: cleavage stimulatory factor; TMG: trimethylated guanosine.

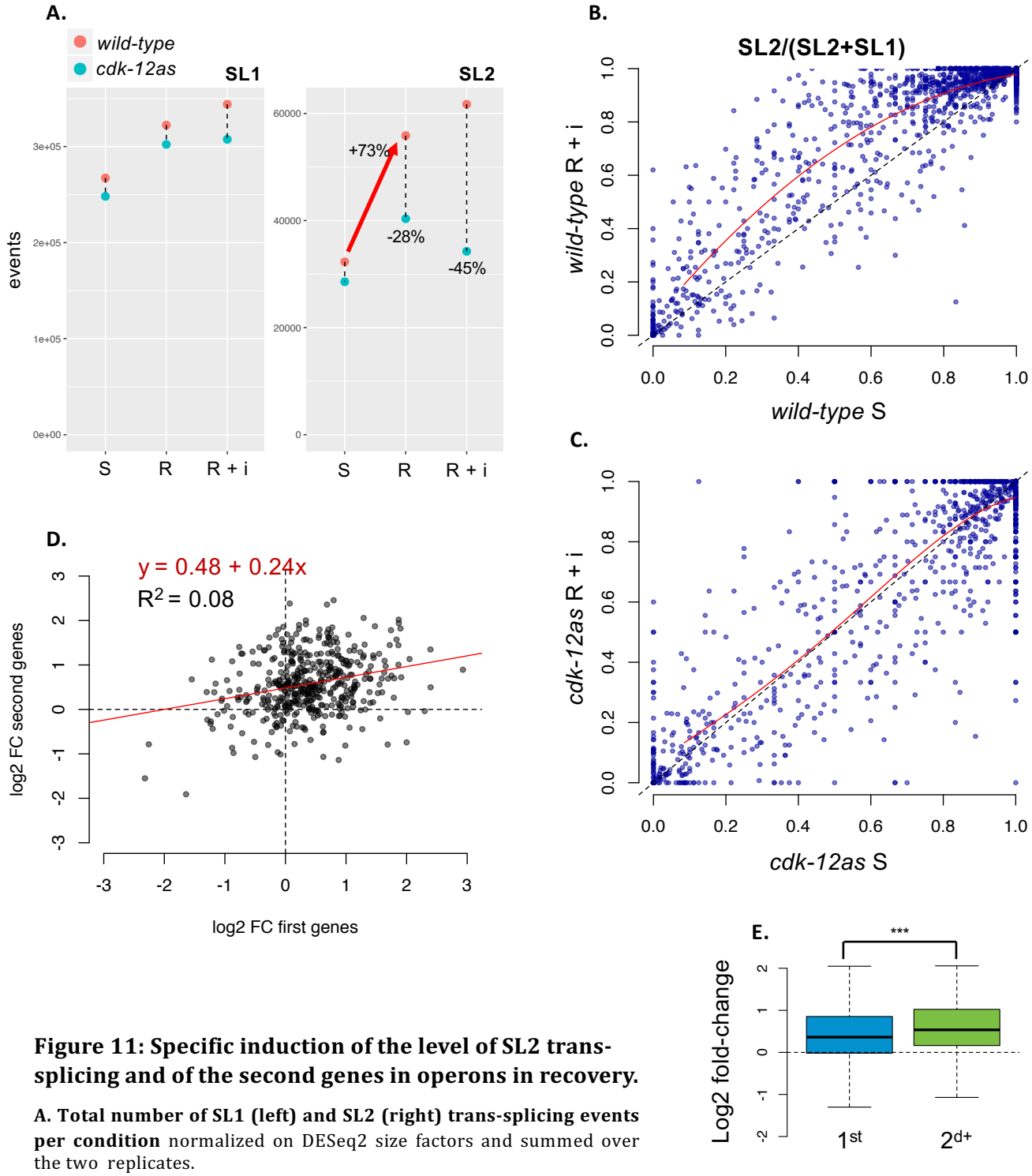


Figure 11: Specific induction of the level of SL2 trans-splicing and of the second genes in operons in recovery.

A. Total number of SL1 (left) and SL2 (right) trans-splicing events per condition normalized on DESeq2 size factors and summed over the two replicates.

B and C. Scatterplot of the proportion of SL2 trans-splicing events for genes in operon in the indicated conditions. A polynomial regression is drawn in red.

D. Scatterplot of the log2 fold change due to the food effect for the second genes in operons compared to the first genes. A linear regression with its equation and coefficient of determination (R^2) are displayed in red.

E. Boxplot of the log2 fold change due to the food effect for the first s(1^{st}) and seconds (2^{d+}) genes in operons. P-value for the paired Wilcoxon-Mann-Whitney tests is of 1.4×10^{-6}.

H. Specific CDK-12-dependent induction of the SL2 trans-splicing upon recovery.

Intriguingly, the global level of SL2-trans-splicing is strongly increased (+73%) in recovery from starvation (**figure 11.A**). This could be simply explained by the fact that genes in operon are globally up-regulated in recovery (Maxwell et al., 2012). However, transcription from the internal promoters (estimated from the level of SL1 trans-splicing) is not induced in recovery, while polycistronic transcription (estimated from the level of SL2 trans-splicing) is significantly induced (**figure 7**).

In consequence for this specific induction, the proportion of SL2 trans-splicing increases for the genes in operons with auxiliary internal promoters (the genes trans-spliced with a mix of SL1 and SL2) (**figure 11.B**). As this switch toward productive polycistronic transcription is CDK-12-dependent (**figure 11.C**), we wondered whether the efficiency of trans-splicing could be used by the cell to finely tune the expression of genes in operons. Consistent with this idea, the expression changes in recovery of the first and second genes in operon are poorly corelated (**figure 11.D**), with the second genes in operons significantly more over-expressed (**figure 11.E**). The existence of internal promoters cannot explain this difference because (1) a minority of operons contains internal promoters with significant activity (**figure 6**), and (2), expression from these internal promoters (as approximated from SL1 trans-splicing quantification) does not generally change in recovery (**figure 7**).

Collectively, our results suggest that the expression of the polycistronic transcripts is uncoupled from the expression of their promoter. This decoupling could be caused by various post-transcriptional processes, such as the modulation of the stability of the mRNAs, but also by co-transcriptional processes, such as the modulation of the efficiency of the SL2 trans-splicing, that, upon failure, causes Pol II to terminate prematurely. This last possibility is plausible given that the efficiency of trans-splicing is dependent on the activity of CDK-12 and that CDK-12 activity itself could be regulated. This last argument needs to be demonstrated in *C. elegans*, but in fission yeast, the activity of the CDK-12 homolog Lsk1 is modulated by the phosphorylation of its N-terminal domain in response to extracellular cue (Materne et al., 2015; Sukegawa et al., 2011).

I. Other roles for CDK-12 activity?

i. Longer genes

All together our results convincingly point toward a model where CDK-12 activity is required for the SL2 trans-splicing. However, many of the protein-coding genes significantly mis-regulated (222 down-regulated and 149 up-regulated) upon CDK-12 inhibition are not in operon. Therefore, their mis-regulation could either be due to undirect effects or reveal other roles for CDK-12 beyond the coordination of the SL2 trans-splicing.

Interestingly, the distribution of the length of the lone genes down-regulated upon CDK-12 inhibition not in operon (with a median gene length of 2.2 Kb) is significantly higher (Wilcoxon-Mann-Whitney unpaired test p-value < 2.2×10^{16}) than for the unaffected genes (with a median gene length of 1.1 Kb) (**figure 12.A**). This result could be biased by the fact that longer genes tend to be more represented in RNA-seq libraries,

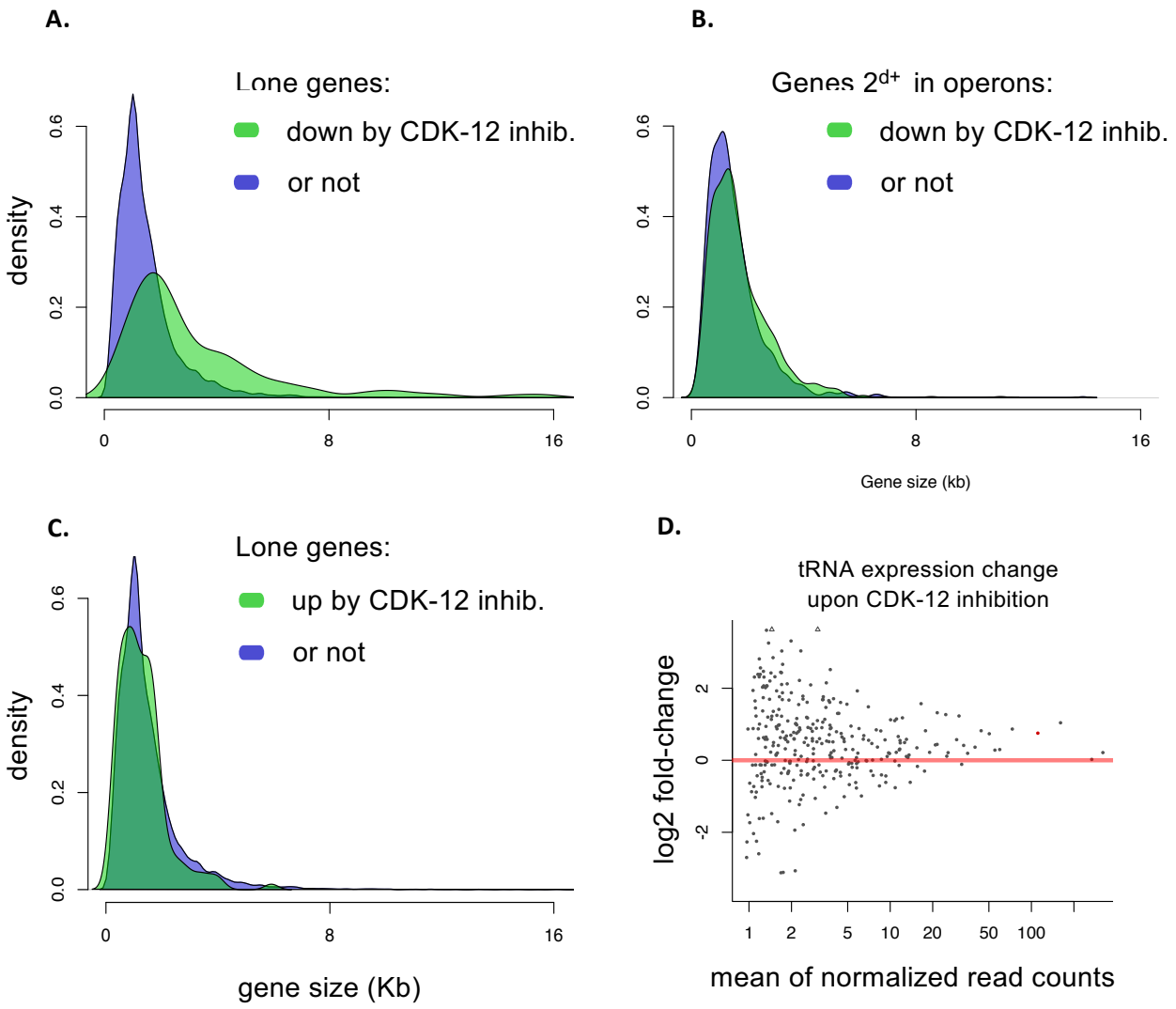


Figure 12: Genes not in operons down-regulated upon CDK-12 inhibition are longer.

A-B-C. Density plot of gene size for the protein-coding genes in the indicated categories. Effectives are as follow: **A.** Lone genes down-regulated by CDK-12 inhibition (n = 222) or not (n = 13754); **B.** genes in position two and over in operons down-regulated by CDK-12 inhibition (n = 765) or not (n = 1246); **C.** Lone genes up-regulated by CDK-12 inhibition (149) or not (13321).

D. MA-plot of the tRNA genes expression change upon CDK-12 inhibition.

providing more read – i.e smore power – for the differential expression analysis. To control for this possible bias, we compared the rather similar gene length distribution for the genes in operon down-regulated or not (**figure 12.B**) and for the lone genes up-regulated or not (**figure 12.C**). Those results convincingly demonstrate that the aforementioned bias is negligible and, in consequence, that CDK-12 activity can be important for the expression of longer genes via a currently unknown mechanism that could become the focus of future studies. Intriguingly, this dependence of longer genes on CDK-12 function is conserved in *Drosophila melanogaster* and human cell lines (Blazek et al., 2011a; Pan et al., 2015).

ii. tRNAs

The studies in fission yeast presented in the first part of this thesis revealed that mutants lacking Ser2-P have increased tRNA expression. Here, no tRNAs were found significantly affected by CDK-12 inhibition in *C. elegans* (**table2**).

However, more careful analysis of the tRNAs expression showed that the number of reads uniquely mapped to the tRNAs is very low and does not provide sufficient power for robust differential expression analysis (**figure 12.D**). Nevertheless, the overall trend tends toward over-expression, which is reminiscent of our findings in fission yeast. While more work is required to corroborate this finding and assess the mechanisms behind it, we also noted that genes annotated as involved in tRNA processing are enriched in the genes down-regulated upon CDK-12 inhibition (Fisher's exact test FDR = 0.014).

Table 2: genes up or down-regulated by CDK-12 inhibition

gene type	up	down
protein-coding	187	1004
lncRNA	15	4
snRNA	1	0
snoRNA	0	0
tRNA	0	0

iii. DNA repair

Studies of CDK-12 function in mammals and drosophila highlighted that CDK-12 is essential for the expression of DNA damage response genes (Blazek et al., 2011a; Li et al., 2016). While our own study is focused on developmental aspect rather than DNA damage, DNA repair genes are enriched in the genes down-regulated upon CDK-12 inhibition (Fisher's exact test FDR = 0.02), suggesting that this function of CDK-12 is conserved in *C. elegans*. Again, more work would be required to confirm that correlation.

3. CONCLUSIONS

i. A gene specific requirement for CDK-12 activity

In this work, we studied the role of Ser2 kinase CDK-12, and, by extension, Ser2-P, in *C. elegans* development. Using a *cdk-12as* (analog-sensitive) strain, we were able to confirm that CDK-12 is dispensable for embryonic development, but is essential for post-L1 development. This stage-specific requirement suggests that CDK-12 is not required for general transcription but that it might be important for the expression of some genes involved in post-L1 development.

Accordingly, we found that only a subset of the genes are sensitive to the inhibition of CDK-12 in a RNA-seq experiment. Consistent with the L1-arrest phenotype of CDK-12 inhibition, the genes affected are enriched in genes overexpressed during the recovery from starvation. Among these genes, many are annotated as being associated with development. Therefore, we propose that the failure to overexpress some of these genes is the causes the requirement for CDK-12 during larval development.

This gene- and stage-specific requirement is reminiscent of an earlier work from our group that describes the specific requirement of Ser2-P for the sexual differentiation and *ste11* expression in fission yeast (Coudreuse et al., 2010). With the reports of similar gene-specific requirements in mammals (Blazek et al., 2011b; Ekumi et al., 2015; Liang et al., 2015) and drosophila (Pan et al., 2015; Li et al., 2016), it appears more and more clearly that Ser2-P is more than just a signal that flag elongating polymerases.

ii. A mechanism for CDK-12 requirement for efficient SL2 trans-splicing

Strikingly, the majority of the CDK-12-dependent genes were found to be organized in operons and were localized more specifically downstream the first genes of the operons. Quantification of the trans-splicing events revealed a defect in SL2-trans-splicing for those genes, highlighting a role for CDK-12 in coupling transcription with the processing of polycistronic transcripts into discrete mRNAs (**figure 10**). A failed or untimely SL2 trans-splicing causes Pol II to terminate early before it reaches the end of the operon, possibly because the uncapped 5' end of the transcript creates an entry point for exonucleases that can terminate Pol II according to the torpedo model (West et al., 2004).

The connection between CDK-12 and SL2 trans-splicing might lie in the described interactions between the termination factor CstF and Ser2-P and between CstF and the SL2 ribonucleoprotein (Garrido-Lecca et al., 2016). However, the fact that the loss of Ser2-P upon CDK-12 inhibition does not globally affect termination *per se* in any tangible way suggest that CstF is still properly recruited, at least at genes not in operons. To clarify this point, a short term perspective would be to assess CstF occupancy upon CDK-12 inhibition at genes outside and within operons.

In conclusion, while the gene-specific requirements for Ser2-P is almost a constant in every organisms where it has been investigated, the mechanism behind it can vary. In fission yeast, we previously described a Ser2-P dependent chromatin-based

mechanism of repression/derepression of *ste11* and other genes (Materne et al., 2015; Materne et al., 2016) while in this study, the gene-specificity lie in the modulation of the SL2 trans-splicing efficiency. As SL trans-splicing has a patchy phylogenetic distribution (Krchnakova et al., 2017), it would be interesting to assess whether the dependence on Ser2-P is also conserved in species evolutionary distant from *C. elegans*, such as trypanosomes (whose Pol II has a non-canonical CTD) or the chordate *Ciona intestinalis*.

iii. Does CDK-12 regulate trans-splicing in response to environmental cues?

Finally, our results show that the expression of the polycistronic transcripts is uncoupled from the expression of their promoter. This decoupling could be caused by various post-transcriptional processes playing on the stability of the mRNAs, but perhaps also by co-transcriptional processes, such as the modulation of the efficiency of the SL2 trans-splicing, that, upon failure, causes Pol II to terminate prematurely. Premature termination would be more economic (in term of cell resources) than post-transcriptional degradation.

As we saw that the efficiency of the SL2 trans-splicing depends on CDK-12 activity, it would be an exciting perspective to assess whether CDK-12 activity is regulated to respond to the presence of food in the environment, therefore connecting the expression of genes in position two and over in operons with extracellular cues. Indeed, condition-specific inductions of CTD kinase activity has been reported in various species as distant as the fission yeast (Materne et al., 2015; Sukegawa et al., 2011) and *Arabidopsis thaliana* (Li et al., 2014) and could therefore be conserved across eukaryotes, bringing new perspectives into the reductive view of CTD phosphorylation that describes it as global, uniform transition from Ser5-P to Ser2-P during the transcription cycle.

MATERIAL & METHODS

1. EXPERIMENTAL PROCEDURES

i. Yeast strains and media

Yeast were grown in YES medium unless stated otherwise. Classical genetic procedures (gene targeting, crosses and growth) were performed as described (Bähler et al., 1998; Forsburg, 2003; Moreno et al., 1991).

The following strains were used, presented in their order of apparition in this thesis:

#	name	from	mating-type	genotype
94	*WT*	Paul Nurse	h-	-
552	*S2A*	(Coudreuse et al., 2010)	h-	*rpb1 CTD S2A -kanR*
1635	*rpc25-flag*	this study	h-	*rpc25-flag-natR*
1576	*rpc25-myc[a]*	(Iwasaki et al., 2010)	h-	*rpc25-myc-kanR ade6-216 ura4-D18 leu1-32 his2*
1590	*rpc25-myc*	this study	h+	*rpc25-myc-kanR*
1577	*rpc25-TAP[a]*	(Iwasaki et al., 2010)	h-	*rpc25-TAP-kanR ade6-216 ura4-D18 leu1-32 his2*
1913	*rpc1-TAP*	François Bachand	h+	*rpc1-TAP-natR ura4-D18 leu1-32 his3-D1*
1914	*rcp2-TAP*	François Bachand	h+	*rpc2-TAP-natR ura4-D18 leu1-32 his3-D1*
1915	*rpc25-TAP*	François Bachand	nd	*rpc25-TAP-natR ura4-D18 leu1-32 his3-D1*
1606	*rpc25-flag S2A*	this study	h-	*rpc25-flag rpb1 CTD S2A -kanR*
581	*Δlsk1*	(Coudreuse et al., 2010)	h+	*lsk1::ura4 ura4-D18*
1861	*Δlsk1*	this study	h+	*rpc25-flag-natR lsk1::ura4 ura4-D18*
1919	*rpc1-TAP*	this study	h-	*rpc1-TAP-natR*
1950	*rpc1-TAP rpc25-flag*	this study	h-	*rp25-flag-hphR rpc1-TAP-natR*
1712	*Δmaf1*	this study	h-	*maf1::natR*
1713	*Δmaf1 S2A*	this study	h-	*maf1::natR rpb1 CTD S2A -kanR*
1850	*Δmaf1 rpc25-flag*	this study	h-	*maf1::natR rpc25-flag-hphR*
1851	*Δmaf1 rpc25-flag S2A*	this study	h-	*maf1::natR rpc25-flag-hphR rpb1 CTD S2A -kanR*
643	*Δelp3*	(Bauer et al., 2012)	h-	*elp3::natR*
1096	*Δste11*	(Materne et al., 2015)	h-	*ste11::kanR*
780	*Δrrp6*	this study	h-	*rrp6::kanR*
1273	*Δdcr1*	this study	h-	*dcr1::natR*
1917	*rpc1-TAP S2A*	this study	h-	*rpb1 CTD S2A -kanR rpc1-TAP-natR*
1719	*brf1-TAP*	this study	h-	*brf1-TAP-kanR*
1720	*brf1-TAP S2A*	this study	h-	*rpb1 CTD S2A -natnR rpc1-TAP-kanR*
1971	*sfc6-TAP*	this study	h-	*sfc6-TAP-kanR*
1972	*scf6-TAP S2A*	this study	h-	*rpb1 CTD S2A -natR sfc6-TAP-kanR*
954	*rsc1-TAP*	(Materne et al., 2015)	h-	*rsc1-TAP-hphR*

1033	*rsc1-TAP S2A*	(Materne et al., 2015)	h-	*rpb1 CTD S2A -kanR rsc1-TAP-hphR*
1873	*rpc25-flag Δrrp6*	this study	h-	*rrp6::kanR rpc25-flag-natR*
1867	*Δcid14*	this study	h-	*cid14::kanR*
1904	*Δcid14 rpc25-flag*	this study	h-	*cid14::kanR rpc25-flag-natR*
1923	*rrp6-flag*	this study	h-	*rrp6-flag-natR*
1916	*rrp6-flag S2A*	this study	h-	*rpb1 CTD S2A -kanR rrp6-flag-natR*

ii. Western blot

Alkaline protein extraction was performed as described (Matsuo et al., 2006). Briefly, 10 mL of yeast culture at OD 0.5 (about $5.3x10^7$ cells) were pelleted, then the pellet was resuspended in water and incubated during 10 minutes in a NaOH 0.3M final solution. After centrifugation, the pellet was resuspended in a buffer containing Tris HCL, SDS, glycerol and β-mercaptoethanol. The extracts where boiled, separated on SDS-PAGE gels (4-15% gradient gels, precast, from Biorad) and transferred to a PVDF membrane using the Trans-Blot Turbo Transfer System (Biorad). The membrane was then blocked in PBS-Tween with 5% milk and incubated with the appropriate antibody for 1 hour. After signal amplification with secondary antibodies covalently bound to horseradish peroxidase, revelation and quantification of the chemiluminescence using ECL was made using an ImageQuant^{TL} apparatus and software. The antibodies used in this study are as follow (also for the ChIP experiments):

target	**ref**
flag	M2 (Sigma F3165)
Pol II Rpb1 CTD	8WG16
Ser2-P	3E10, ab5095
H3	AB1791.1
TAP	Sigma P1291
Tubuline	Sigma T5168

iii. Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation experiments were performed as described in (Materne et al., 2015; Migeot and Hermand, 2018).

Briefly, for the chromatin extraction, 80 mL of cultures at OD 0.7 were crosslinked with 1% formaldehyde (final concentration) during 10 minutes. Then, the crosslinking reaction was quenched using excess glycine (final concentration of 0.4M) during 5 minutes. From this point, we worked on ice or in the 4°C room. The crosslinked cells where pelleted, washed in a Tris HCL 0.02M solution, then resuspended in a detergent-containing (0.1% SDS, 1% tritonX100) FA/SDS solution supplemented with protease inhibitors (PMSF). Then, the crosslinked cells were lysed by 7 cycles of 20 seconds in the *fast-prep* apparatus, which physically shear the cells by vigorous shaking in presence of

zirconium beads. The chromatin was then purified from the cell extracts by two cycles of high-speed centrifugations (the pellet contains the chromatin) - resuspension in FA/SDS/PMSF. Then the chromatin was sheared by sonication in the Bioruptor (Diagenode) apparatus (7 cycles of 30 seconds ON, 1 minutes OFF, in high voltage mode), providing chromatin fragments of about 300 bp.

For the chromatin immunoprecipitation, the sheared chromatin was incubated during 2 hours at 21°C with magnetic beads (Diagenode) coupled to the antibody of interest (see previous table in the *western-blot* section). After multiple washes in high salt (0.5M NaCl) FA/SDS buffer at room temperature, the chromatin bound to the magnetic beads was eluted by 20 minutes incubation at 65°C. Then, the proteins from the immunoprecipitated chromatin extracts were digested by proteinase K during 1h at 37°C. After overnight incubation at 65°C to decrosslink the chromatin, RNA contamination was eliminated with RNaseA treatment (1h at 37°C). Finally, the DNA from the immunoprecipitated chromatin was further purified on *rapace* PCR purification columns and resuspended in water.

Finally, enrichment at target loci was assessed by qPCR (*quantitative Polymerase Chain Reaction*), by comparing SyberGreen (Roche) fluorescent signal between the immunoprecipitated samples and the input samples[11]. The following oligonucleotides were used for PCR amplification (F = forward, R= reverse), presented in their order of apparition in this thesis:

name	**sequence**
ARG.05 F	AACCAGGCAAAGGTTGTTAC
ARG.05 R	TAATTCACTCCCCAACAACG
ILE.04 F	TGGCAAGAGTGGTGTCCATC
ILE.04 R	AACCGACTACATCATGCGAC
TYR.04 F	CAAGCACCGGCTATACAACAC
TYR.04 R	TGGAAGAGAGCTTGCCTTAGTG
snU6 (+192) F	TTCCCATGTTGTCTCCAACC
snU6 (+192) R	TTCCCATGTTGTCTCCAACC
ILE.04 +1000 F	TTGAGTCGAACTTAGCAAATGG
ILE.04 +1000 R	GTTGCGAAACATAGCCTCTTAC
adh1 promoter (-132) F	TTCCCATGTTGTCTCCAACC
adh1 promoter (-132) R	TTCCCATGTTGTCTCCAACC
srp7 F	GGGTTTGCAATGAAAAGTTGA
srp7 R	ACTAAACACTTCGACCAAGC
ste11 -2842 F	TTCCCATGTTGTCTCCAACC
ste11 -2842 R	TTCCCATGTTGTCTCCAACC

[11] The input samples come from the same chromatin extract than their immunoprecipitated counterparts and undergo same processing steps (proteinase K, decrosslink, RNAse A, purification), but without the immunoprecipitation.

ste11 -2465	TTCCCATGTTGTCTCCAACC
ste11 -2465	TTCCCATGTTGTCTCCAACC
ste11 -1337	TTCCCATGTTGTCTCCAACC
ste11 -1337	TTCCCATGTTGTCTCCAACC
snU6 +379 F	CAAATCTTGATTTCCAATCGCT
snU6 +379 R	TTTGCGACTTAGCAAGAAGT
act1 F	CCACTATGTATCCCGGTATTGC
act1 R	CAATCTTGACCTTCATGGAGCT
5S rRNA F	TAGGCGAAAACACCAGTTCC
5S rRNA R	TTCCCATGTTGTCTCCAACC
PRO.09 F	GCCGTTTGGTCTAGTGGTATG
PRO.09 R	TTGGGCTGTTGTGGGAATC

iv. Northern-blot and strand-specific RT-qPCR

Total RNA was prepared using the classical hot phenol protocol, where after cell lysis, three successive phenol extraction followed by one ethanol precipitation and two ethanol washes are applied to ensure high RNA purity (Schmitt et al., 1990). For the tRNA Northern blots, 8 µg of total RNA extracts were loaded on 10% acrylamide-urea gels as described (Wu et al., 2015) and detection was made using the following ^{32}P-labeled oligonucleotides:

name	sequence
ARG.05	TAGGAATGAGATGCGCTACCATTGCGCCA
LEU.CAA	GTATTACTTGAGTGCTGCGCCATAGACCGC
5S rRNA	TTCCCATGTTGTCTCCAACC

For the strand-specific RT-qPCR, complementary DNA (cDNA) was synthetized from the total RNA extract using the SuperScript II reverse in presence of actinomycin to enforce strand-specificity.

After purification of the cDNA products, a quantitative PCR reaction allowed to infer the relative quantity of RNA in the samples using the ΔΔCt method on the SyberGreen (Roche) fluorescent signal.

v. RNA-seq library preparation

RNA-seq library preparation was made using the Illumina TruSeq stranded total RNA library preparation kit following manufacturer's instructions, with the notable exception that the quantity of AMPure XP beads used for clean ups was scaled up by a ratio of 1.1 in order to retrieve smaller RNA species such as tRNAs.

vi. *ChIP-seq and genomic DNA library preparations*

ChIP extracts were prepared as for the ChIP-qPCR experiments. Total purification of genomic DNA was made using Zymo research Kit for genomic DNA extraction (YeaStarTM Genomic DNA Kit) then the DNA was fragmented by sonication using the Bioruptor (Diagenode) set on Low position (130 W) for 7 cycles of 30 seconds. A fraction of the sonicated DNA was charged on a 1 % agarose gel in order to check the fragment size. The resulting fragments were sequenced, along with ChIP extracts, using the Illumina TruSeq ChIP-seq library preparation protocol according to manufacturer's instructions.

2. BIOINFORMATICS ANALYSIS

i. *Generalities: quality control & figure drawing.*

For all high throughput sequencing experiments, the reads quality was assessed using FastQC. Quality and adaptor clipping was performed using trimmomatic (Bolger et al., 2014). Statistical analysis and figure drawing was performed within the R programming environment (R Core Team, 2017).

ii. *MNase-seq data analysis*

The MNase-seq data was analyzed as described in the section "bioinformatics analysis" of the Appendix I.

iii. *RNA-seq analysis*

Reads were mapped on the *S. pombe* genome or the *C. elegans* genome using hisat2 (Kim et al., 2015) as described in the Appendix II. The number of reads by region of interest was quantified using featureCounts (Liao et al., 2014). Differential expression analysis was carried on with DESeq2 (Love et al., 2014). Criteria for a differentially expressed feature are a false discovery rate (FDR) < 0.01 and an absolute fold change > 1.5.

The quantification of trans-splicing events from the RNA-seq data was performed using SL-quant (thoroughly described in Appendix II) in sensitive mode.

iv. *ChIP-seq analysis*

Reads were mapped on the *S. pombe/S. cerevisiae* genomes or the *C. elegans* genome using bowtie2 (Langmead and Salzberg, 2012). Normalization of the signal tracks for the *S. pombe* experiments was either made by scaling on the total number of reads mapped to the *S. pombe* nuclear chromosomes (excluding the rDNA repeats), or to the number of spike-in reads mapped on *S. cerevisiae* genome for the IPed samples divided by

the proportion of spike-in reads in the corresponding input samples. The *C. elegans* profiles were normalized using the BEADS algorithm (Cheung et al., 2011).

APPENDIX I

Yague-Sanz, C., Vazquez, E., Sanchez, M., Antequera, F., and Hermand, D.

A conserved role of the RSC chromatin remodeler in the establishment of nucleosome-depleted regions.

2017, Current Genetics, *63*, 187-193.

While we were studying the Ser2-P- and RSC-dependent nucleosome dynamics that occur at the promoter of *ste11* and of other genes in fission yeast by MNase-seq (Materne et al., 2015; Materne et al., 2016), another group published MNase-seq data in budding yeast in a RSC-depleted strain (Parnell et al., 2015). We seized that opportunity to compare the changes in nucleosome occupancy following RSC impairment (by depletion or deletion of a non-essential subunit) in the two distantly related yeast species, and concluded, in contrast with earlier studies (Pointner et al., 2012), that the RSC complex is required, at least partially, for NDR formation in both species.

ABSTRACT

The occupancy of nucleosomes governs access to the eukaryotic genomes and results from a combination of biophysical features and the effect of ATP-dependent remodelling complexes. Most promoter regions show a conserved pattern characterized by a nucleosome-depleted region (NDR) flanked by nucleosomal arrays. The conserved RSC remodeler was reported to be critical to establish NDR *in vivo* in budding yeast but other evidences suggested that this activity may not be conserved in fission yeast. By reanalysing and expanding previously published data, we propose that NDR formation requires, at least partially, RSC in both yeast species. We also discuss the most prominent biological role of RSC and the possibility that non-essential subunits do not define alternate versions of the complex.

NOTE: Due to copyright restrictions, the formated version of this article could not be included in this thesis.

Go to https://doi.org/10.1007/s00294-016-0642-y to view the fully formated version (in open access).

A conserved role of the RSC chromatin remodeler in the establishment of nucleosome depleted regions

Carlo Yague-Sanz[1], Enrique Vázquez[2], Mar Sánchez[2], Francisco Antequera[2] and Damien Hermand[1,#]

[1] URPHYM-GEMO, Namur Research College (NARC), The University of Namur, Namur 5000 Belgium.

[2] Instituto de Biología Funcional y Genómica, Consejo Superior de Investigaciones Científicas (CSIC)/Universidad de Salamanca, Campus Miguel de Unamuno, 37007 Salamanca, Spain.

*Correspondence to: *Damien.Hermand@unamur.be*
Keywords: yeast, chromatin, RSC, nucleosome, mitosis

Abstract

The occupancy of nucleosomes governs access to the eukaryotic genomes and results from a combination of biophysical features and the effect of ATP-dependent remodeling complexes. Most promoter regions show a conserved pattern characterized by a nucleosome-depleted region (NDR) flanked by nucleosomal arrays. The conserved RSC remodeler was reported to be critical to establish NDR *in vivo* in budding yeast but other evidences suggested that this activity may not be conserved in fission yeast. By reanalysing and expanding previously published data, we propose that NDR formation requires, at least partially, RSC in both yeast species. We also discuss the most prominent biological role of RSC and the possibility that non-essential subunits do not define alternate versions of the complex.

Introduction

A fundamentally different logic of gene regulation between procaryotes and eukaryotes was previously proposed based on the existence of chromatin in the latter, which results in a closed, less accessible genome (Struhl, 1999). Despite the fact that this view may be too simplistic - there are structural proteins associated with the DNA in procaryotes (Anuchin et al., 2011) and eukaryotes use more repressors than anticipated (Kemmeren et al., 2014)- the general concept still stands true (Estrada et al., 2016). Strikingly, the human genome harbours the blueprint of about 200 highly specialized cell types characterized by very different morphology, metabolism and capacities including for example neurons, hepatocytes or gametes. By comparison, most procaryotes have a limited range of cellular states. Therefore, the invention of chromatin may have been pivotal for the emergence of highly differentiated cell types, most likely because the expression of specific programmes must be tightly regulated to allow diverse and sometimes antagonistic differentiated states to co-exist. For example, yeast differentiation during gametogenesis must be very strictly limited to diploid cells to avoid massive cell death resulting from haploid meiosis and recent data support that chromatin-based mechanisms play a key role in that process (van Werven et al., 2012). The understanding of how chromatin is established and how it contributes with most, if not all nuclear processes including transcription, DNA replication, DNA repair, recombination or chromosome segregation therefore constitutes an outstanding focus in current biology. In that context, an important

and long-standing question is to decipher how nucleosomes, which constitutes the basic unit of chromatin, are positioned genome-wide. There has been abundant debates about what dictates the position of nucleosomes with models fully relying on biophysics - the position is DNA encoded – and a model encompassing a layer of active, ATP-dependent modelling of the chromatin template. The reader is redirected to excellent reviews addressing these issues (Korber, 2012; Lieleg et al., 2015; Struhl and Segal, 2013). In the meantime, the discovery of a large set of chromatin remodelers supported the second possibility and the development of more refined nucleosome occupancy maps in various species led to the discovery of general features of nucleosome positioning along the eukaryotic genomes. Particularly, nearby the promoter, there is often a stereotypical organization just upstream of the transcription start site (TSS) with a large nucleosome depleted region (NDR) flanked by highly positioned nucleosomes referred to as the +1 and the -1. From these two landmarks, arrays of regularly-spaced nucleosomes often extend. The RSC (Remodel the Structure of Chromatin) complex has a specific role in the generation of NDR in budding yeast (Badis et al., 2008; Hartley and Madhani, 2009; Parnell et al., 2008; Wippo et al., 2011). RSC is an abundant and essential paralog of the canonical SWI/SNF remodeler (Cairns et al., 1996) that contains a DNA-dependent ATPase (Sth1 in budding yeast, Snf21 in fission yeast) that translocates DNA and allows to shift nucleosome positions, or even completely eject nucleosomes. Importantly, RSC show compositional and functional differences between budding and fission yeasts (Monahan et al., 2008) and it was suggested that RSC is not required for NDR formation in the latter (Pointner et al., 2012). In addition, not all subunit of RSC are essential, which may indicate that subcomplexes exist and may have specialized functions. Finally, it is unclear what is the most prominent biological role of RSC, in other words which nuclear process is mainly affected upon RSC inactivation and results in lethality. Here we briefly comment on these issues by reanalysing previously published data from budding yeast and by building on our recent work in fission yeast.

Defects in the RSC remodeler affects NDR formation in both budding yeast and fission yeast

One of the early difficulties in comparing the two model yeast species (the budding yeast *Saccharromyces cerevisae*, hereafter *S. cerevisiae,* and the fission yeast *Schizosaccharomyces pombe*, hereafter *S. pombe*) was the very poor annotation of the transcription start sites in *S. pombe*. This is an issue when plotting the average nucleosome signal obtained from MNase-Seq experiments to transcripts coordinates. Very recently, several groups (Booth et al., 2016; Eser et al., 2016; Li et al., 2015) have used different approaches to address that issue and **Figure 1A** shows a comparison of the average nucleosome signal when plotted to the Pombase annotation (the reference of the community working on fission yeast) and the most recent work performed by the Lis laboratory using Precision Run-On 5' cap sequencing (PRO-cap) (Booth et al., 2016), which corresponds to the annotation used in this manuscript. Comparing the overlay of MNase-Seq data after alignment at the TSS between the two yeast species confirms previous observations from the pioneer work of the Korber laboratory in fission yeast, namely the absence of a clearly positioned -1 nucleosome and shorter nucleosome spacing in *S. pombe* (Lantermann et al., 2009). In addition, the +1 nucleosome is positioned further away from the TSS and there are clear, albeit weak nucleosome arrays upstream of NDR in fission yeast (**Figure 1B**). Importantly, the low amplitude of the peaks and the absence of a positioned -1 nucleosome were previously shown to result from the larger variation in the size of individual NDRs in fission yeast (Soriano et al., 2013). Therefore, nucleosomal arrays emanate bidirectionally from the NDRs in fission yeast as well.

It is now well established that the ablation of RSC activity in budding yeast strongly affects the majority of NDRs (Badis et al., 2008; Hartley and Madhani, 2009;

Parnell et al., 2008). Recent data further confirmed that in the inactivation of RSC (resulting from switching-off the expression of *rsc8*) leads to upstream and downstream nucleosomal arrays to shift to and eventually occlude the NDR. This work also supports the idea that phasing patterns reflect the resultant of phasing signals emanating from neighbouring NDRs (Ganguli et al., 2014). Most recently, it was reported by the Cairns laboratory that RSC and ISW1 have functional antagonism, which is supported by the fact that the gain in nucleosome occupancy in *rsc* mutant is attenuated by the additional inactivation of ISW1 (Parnell et al., 2015).
In the fission yeast, the RSC complex was also associated with the generation of NDR in the context of heterochromatin. Indeed the deletion of *rsc1* that encodes a non-essential subunit of RSC suppresses the requirement of the histone deacetylase (HDAC) Clr3 for NDR elimination (Garcia et al., 2010), which indirectly supports that RSC is responsible for acetylation-dependent NDR formation in that species. However, that study did not expand to euchromatin. Finally, it was reported that CHD1 remodelers, Hrp1 and Hrp3 are required in fission yeast to link nucleosomal arrays to most TSS (Pointner et al., 2012). In the same study, the role of RSC in nucleosome positioning was also analysed using the only conditional (thermosensitive) mutant allele available for the gene *snf21* that encodes the catalytic subunit of RSC (Yamada et al., 2008). Unexpectedly, no effect on nucleosome positioning around TSS was obvious upon thermal inactivation of the mutant, which suggested that RSC plays no role in NDR formation in *S. pombe*.
Our recent work identified RSC as a key downstream effector of a cascade controlling the level of acetylation around the NDR of *ste11*, which encodes the master regulator of gametogenesis in *S. pombe* (Anandhakumar et al., 2013; Cassart et al., 2012; Coudreuse et al., 2010; Devos et al., 2015). Deletion of non-essential subunits of RSC or transcriptional switch-off of *snf21* both impede *ste11* expression and correlate with higher nucleosome occupancy at the *ste11* NDR (Materne et al., 2015; Materne et al., 2016). This effect led us to analyse the genome-wide effect of both RSC mutants on nucleosome positioning using MNase-Seq, which allows us to assess the conservation of the chromatin remodelling function of the RSC complex in budding and fission yeast.
We have reanalysed the data presented in the Parnell et al. paper (Parnell et al., 2015) with the following modifications compared to the published work. First, no filtering for specific gene organization was applied. Second, the data are presented at single base pair resolution rather that within 50bp windows relative to the TSS as done before. This was made possible as we used the MNase-Seq data rather than the Agilent 244K mircoarrays used for the main figures of the original study. Third, the TSS annotation was obtained from a different source (Ganguli et al., 2014). **Figure 2A** shows the profile of nucleosome occupancy ratios between the *sth1* degron (a strain that allows rapid and conditional degradation of the Sth1 protein RSC subunit) and control strains +/- 750 bp relative to the TSS. Genes were organized into clusters based on their *rsc*-/RSC+ ratio. As reported in the published work, a leftward shift in nucleosome positions over the transcribed region is observed, confirming that nucleosomal arrays emanate from the NDR flanking nucleosomes. Most likely thus, NDR filling reflects encroachment by the flanking -1 and +1 nucleosomes, rather than insertion of an additional nucleosome within the NDR region as previously discussed (Ganguli et al., 2014).
We next applied identical analyses to the data obtained in a fission yeast *rsc1* mutant. It should be noted straight away that Rsc1 is a non-essential subunit of RSC while the work done in budding yeast targeted the gene encoding the catalytic subunit. Nonetheless, the general picture obtained when *rsc1* is absent (**Figure 2B**) is reminiscent of the budding yeast data (**Figure 2A**). Compared to budding yeast, the shortening of NDR is observed with a slight shift towards the TSS that recalls the shift of the position of the +1 nucleosome between budding and fission yeast (**Figure 1B**). In addition, the leftward shift in nucleosome positions over the transcribed region is also obvious in most clusters. These data suggest that similarly to budding yeast, RSC also play an important conserved role in the establishment of NDRs in fission yeast in contrast to previous conclusions (Pointner et al., 2012). However, it is

important to keep in mind that the previous study relies on a *ts* allele of *snf21*. As rightly pointed by the authors in their manuscript, it is possible that the inactivation of the *snf21-ts* was not complete despite the fact that the strain has obvious phenotypes (see below), somehow masking an effect of most NDR. Supporting this possibility, we report here that while switching-off *snf21* expression using a Tet-off system eventually results in cell death on plates (Materne et al., 2015), it has a weak genome-wide effect on NDR formation when a short time point is used (**Figure 3A**). This suggests that lowering RSC activity by switching-off the transcription of the *snf21* gene (with about 35% of the *snf21* mRNA left) may only affect the most sensitive RSC-dependent processes. These data claim for the generation of a much more efficient switch-off system that could quickly deplete the vast majority of the Snf21 protein pool in the cell.

A single RSC complex containing non-essential subunits likely fulfils all RSC functions

Interestingly, the analysis of *rsc1* deletion in fission yeast reveals a genome-wide effect on NDR, yet the effect appears milder than the inactivation of the catalytic subunit in *S. cerevisiae* (compare **Figure 2A** and **2B**). However, the shrinkage of NDR when *rsc1* is deleted is statistically significant (one-sample Wilcoxon test pval < 0.01), **Figure 3B**). To us, these data do not support that the non-essential subunits constitute an RSC submodule with specific functions but rather indicates that within RSC, the non-essential subunits may have a less prominent structural role than essential subunits. In line with this, all the phenotypes reported when inactivating RSC, including chromosomal segregation defects, and sensitivity to drugs, are shared to various degrees by all mutants (Monahan et al., 2008).

A key role of RSC in mitotic chromosome condensation

Despite the genome-wide defect observed in NDR formation (**Figure 2B**), the *rsc1* mutant has a subtle impact on steady-state transcription. Expression alteration effects in either direction were seen for only 1.4% of *S pombe* genes in the *rsc1* mutant and other non-essential subunit similarly affect the transcriptome (Monahan et al., 2008). Interestingly, the third cluster in **Figure 2B** harbours the highest frequency of genes affected in the *rsc1* mutant (**Figure 3C**), including *ste11* (Materne et al., 2015; Materne et al., 2016). That cluster is characterized by an increase in occupancy over a broad region upstream of the TSS that may be typical of highly regulated genes relying on larger regulatory sequences, as typically seen for *ste11* (Anandhakumar et al., 2013). Nonetheless, the main phenotypes of the *rsc1* mutant, and the *snf21-ts* mutant for that matter, are cell elongation associated with chromosomes segregation defects. Although these phenotypes may result from the reduced expression of specific genes, a recent study rather points to a direct role of RSC and nucleosome eviction in condensin loading and chromosome condensation (Toselli-Mollereau et al., 2016). A genetic screen for functional partners of condensin in fission yeast (synthetic lethality with *cut3-477* that encodes a condensin ATPase subunit) identified alleles of *arp9* and *snf21* that both showed high frequency of chromatin bridges in anaphase (Robellet et al., 2014). Further analyses revealed the preferred localization of condensin at, or near NDR and that increased nucleosome occupancy upon RSC downregulation is sufficient to decrease condensin binding. These data point to a prominent role of RSC is establishing the landscape of condensin binding during mitosis by establishing NDR.

In conclusion it appears that the role of RSC in NDR formation is conserved between both *S. pombe* and *S. cerevisae*. The milder effect observed with the *rsc1* mutant in fission yeast may be caused by (1) the fact that this subunit is not essential and its deletion could only partially impair RSC activity; (2) the presence of an additional remodeler involved in NDR formation; (3) the fact that *S. pombe* lacks ISW1-type remodelers that are known to oppose the action of RSC in NDR formation in budding yeast (Parnell et al.). Future work will clarify this issue. Biologically, the most prominent role of RSC may be to maintain proper chromosome segregation and

may extend to kinetochore function, sister chromatid cohesion and DNA repair, in addition to its role in promoting transcription (Cao et al., 1997; Hsu et al., 2003; Huang et al., 2004; Shim et al., 2007).

Bioinformatic analyses.

TSS annotations for *S. pombe* were obtained from pombase (http://www.pombase.org/, ASM294v2.26) and (Booth et al., 2016). TSS annotation for *S.cerevisiae* was obtained from (Ganguli et al., 2014). Nucleosome genome-wide occupancy profiles at 1bp resolution were generated as described (Bauer and Hermand, 2012; Drogat et al., 2012; Lenglez et al., 2010; Materne et al., 2015) using DANPOS for the *S. pombe* data or directly downloaded from GEO (supplementary of GSE65593) for the *S. cerevisiae* data (Parnell et al., 2015).

Nucleosome occupancy ratios were computed as the log2 ratio between treatment and control profiles centered at TSS for all protein coding genes. A pseudocount of +1 was added to both the numerator and the denominator to avoid division by zero. Clustering of the nucleosome occupancy ratio was made using the *kmeans()* function in base R with default parameters and k=5 and visualized with the *heatmap.2()* function from the "gplots" package.

Nucleosome positions were computed for each dataset as the local maximum of nucleosome occupancy in a 100 bp window using the *localMaximum()* function from the "MassSpecWavelet" package. Promoter NDR length is computed as the distance in bases between the position of the first nucleosome before the TSS and the first nucleosome after the TSS. This metric was computed for mutants and wild type strains and substracted accordingly to obtain the NDR length difference as in Figure 3B.

Data access

The nucleosome sequencing data are available in the GEO database under the accession numbers GSE84912 (*S. pombe* datasets) and GSE65593 (*S. cerevisiae* datasets).

Compliance with Ethical Standards

The authors declare that they have no conflict of interest.

Ethical approval: this article does not contain any studies with human participants or animals performed by any of the authors.

Acknowledgments

We thank Philipp Korber for critical reading of the manuscript. This work was supported by grant BFU2014-52143-P from the Spanish Ministerio de Economía y Competitividad to FA and by grants PR T.0012.14 to DH. CY is a FRIA Research Fellow. DH is a FNRS Senior Research Associate.

References

Anandhakumar, J., Fauquenoy, S., Materne, P., Migeot, V., and Hermand, D. (2013). Regulation of entry into gametogenesis by Ste11: the endless game. Biochem Soc Trans *41*, 1673-1678.

Anuchin, A.M., Goncharenko, A.V., Demidenok, O.I., and Kaprel'iants, A.S. (2011). [Histone-like proteins of bacteria (review)]. Prikl Biokhim Mikrobiol *47*, 635-641.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., *et al.* (2008). A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell *32*, 878-887.

Bauer, F., and Hermand, D. (2012). A coordinated codon-dependent regulation of translation by Elongator. Cell Cycle *11*, 4524-4529.

Booth, G.T., Wang, I.X., Cheung, V.G., and Lis, J.T. (2016). Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. Genome Res *26*, 799-811.

Cairns, B.R., Lorch, Y., Li, Y., Zhang, M., Lacomis, L., Erdjument-Bromage, H., Tempst, P., Du, J., Laurent, B., and Kornberg, R.D. (1996). RSC, an essential, abundant chromatin-remodeling complex. Cell *87*, 1249-1260.

Cao, Y., Cairns, B.R., Kornberg, R.D., and Laurent, B.C. (1997). Sfh1p, a component of a novel chromatin-remodeling complex, is required for cell cycle progression. Mol Cell Biol *17*, 3323-3334.

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

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

Devos, M., Mommaerts, E., Migeot, V., van Bakel, H., and Hermand, D. (2015). Fission yeast Cdk7 controls gene expression through both its CAK and C-terminal domain kinase activities. Mol Cell Biol *35*, 1480-1490.

Drogat, J., Migeot, V., Mommaerts, E., Mullier, C., Dieu, M., van Bakel, H., and Hermand, D. (2012). Cdk11-cyclinL controls the assembly of the RNA polymerase II mediator complex. Cell Rep *2*, 1068-1076.

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

Estrada, J., Wong, F., DePace, A., and Gunawardena, J. (2016). Information Integration and Energy Expenditure in Gene Regulation. Cell *166*, 234-244.

Ganguli, D., Chereji, R.V., Iben, J.R., Cole, H.A., and Clark, D.J. (2014). RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. Genome Res *24*, 1637-1649.

Garcia, J.F., Dumesic, P.A., Hartley, P.D., El-Samad, H., and Madhani, H.D. (2010). Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. Genes Dev *24*, 1758-1771.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445-458.

Hsu, J.M., Huang, J., Meluh, P.B., and Laurent, B.C. (2003). The yeast RSC chromatin-remodeling complex is required for kinetochore function in chromosome segregation. Mol Cell Biol *23*, 3202-3215.

Huang, J., Hsu, J.M., and Laurent, B.C. (2004). The RSC nucleosome-remodeling complex is required for Cohesin's association with chromosome arms. Mol Cell *13*, 739-750.

Kemmeren, P., Sameith, K., van de Pasch, L.A., Benschop, J.J., Lenstra, T.L., Margaritis, T., O'Duibhir, E., Apweiler, E., van Wageningen, S., Ko, C.W., *et al.* (2014). Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. Cell *157*, 740-752.

Korber, P. (2012). Active nucleosome positioning beyond intrinsic biophysics is revealed by in vitro reconstitution. Biochem Soc Trans *40*, 377-382.

Lantermann, A.B., Straub, T., Stralfors, A., Yuan, G.C., Ekwall, K., and Korber, P. (2009). Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nat Struct Mol Biol *17*, 251-257.

Lenglez, S., Hermand, D., and Decottignies, A. (2010). Genome-wide mapping of nuclear mitochondrial DNA sequences links DNA replication origins to chromosomal double-strand break formation in Schizosaccharomyces pombe. Genome Res *20*, 1250-1261.

Li, H., Hou, J., Bai, L., Hu, C., Tong, P., Kang, Y., Zhao, X., and Shao, Z. (2015). Genome-wide analysis of core promoter structures in Schizosaccharomyces pombe with DeepCAGE. RNA Biol *12*, 525-537.

Lieleg, C., Krietenstein, N., Walker, M., and Korber, P. (2015). Nucleosome positioning in yeasts: methods, maps, and mechanisms. Chromosoma *124*, 131-151.

Materne, P., Anandhakumar, J., Migeot, V., Soriano, I., Yague-Sanz, C., Hidalgo, E., Mignion, C., Quintales, L., Antequera, F., and Hermand, D. (2015). Promoter nucleosome dynamics regulated by signaling through the CTD code. eLife *4*.

Materne, P., Vazquez, E., Sanchez, M., Yague-Sanz, C., Anandhakumar, J., Migeot, V., Antequera, F., and Hermand, D. (2016). Histone H2B ubiquitylation represses gametogenesis by opposing RSC-dependent chromatin remodeling at the ste11 master regulator locus. eLife *5*.

Monahan, B.J., Villen, J., Marguerat, S., Bahler, J., Gygi, S.P., and Winston, F. (2008). Fission yeast SWI/SNF and RSC complexes show compositional and functional differences from budding yeast. Nat Struct Mol Biol *15*, 873-880.

Parnell, T.J., Huff, J.T., and Cairns, B.R. (2008). RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. EMBO J *27*, 100-110.

Parnell, T.J., Schlichter, A., Wilson, B.G., and Cairns, B.R. (2015). The chromatin remodelers RSC and ISW1 display functional and chromatin-based promoter antagonism. eLife *4*, e06073.

Pointner, J., Persson, J., Prasad, P., Norman-Axelsson, U., Stralfors, A., Khorosjutina, O., Krietenstein, N., Svensson, J.P., Ekwall, K., and Korber, P. (2012). CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. EMBO J *31*, 4388-4403.

Robellet, X., Fauque, L., Legros, P., Mollereau, E., Janczarski, S., Parrinello, H., Desvignes, J.P., Thevenin, M., and Bernard, P. (2014). A genetic screen for functional partners of condensin in fission yeast. G3 (Bethesda) *4*, 373-381.

Shim, E.Y., Hong, S.J., Oum, J.H., Yanez, Y., Zhang, Y., and Lee, S.E. (2007). RSC mobilizes nucleosomes to improve accessibility of repair machinery to the damaged chromatin. Mol Cell Biol *27*, 1602-1613.

Soriano, I., Quintales, L., and Antequera, F. (2013). Clustered regulatory elements at nucleosome-depleted regions punctuate a constant nucleosomal landscape in Schizosaccharomyces pombe. BMC Genomics *14*, 813.

Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell *98*, 1-4.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat Struct Mol Biol *20*, 267-273.

Toselli-Mollereau, E., Robellet, X., Fauque, L., Lemaire, S., Schiklenk, C., Klein, C., Hocquet, C., Legros, P., N'Guyen, L., Mouillard, L., *et al*. (2016). Nucleosome eviction in mitosis assists condensin loading and chromosome condensation. EMBO J.

van Werven, F.J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., and Amon, A. (2012). Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell *150*, 1170-1181.

Wippo, C.J., Israel, L., Watanabe, S., Hochheimer, A., Peterson, C.L., and Korber, P. (2011). The RSC chromatin remodelling enzyme has a unique role in directing the accurate positioning of nucleosomes. EMBO J *30*, 1277-1288.

Yamada, K., Hirota, K., Mizuno, K., Shibata, T., and Ohta, K. (2008). Essential roles of Snf21, a Swi2/Snf2 family chromatin remodeler, in fission yeast mitosis. Genes & genetic systems *83*, 361-372.

Figures legends

Figure 1: Average nucleosome occupancy nearby the TSS in fission and budding yeasts

A. Average nucleosome signals centered on two different fission yeast TSS annotations.

B. Average nucleosome signals centered on TSS in fission yeast and budding yeasts.

Figure 2: Defects in the RSC remodeler affects NDR formation in both budding yeast and fission yeast

A. The profile of nucleosome occupancy ratios between budding yeast *sth1* degron (*sth1*^{deg}) and control strains is presented as an heatmap, where blue represents a gain in nucleosome occupancy and red represents a loss within a region ranging from -750 bp to +750 bp around the TSS at single nucleotide resolution. Rows represent genes and are organized into 5 groups by k-means clustering.

B. Same as in A, except that the fission yeast *rsc1* deletion mutant (*rsc1*□) and control strains are presented.

Figure 3: Comparison of the fission yeast *rsc1* deletion mutant and the *snf21* switch-off strains

A. The profile of nucleosome occupancy ratios between fission yeast *rsc1* deletion mutant (*rsc1*□) and control strains on the left panel (note that this panel is identical to **Figure 2B** and repeated here for clarity), and between the fission yeast *snf21* switch-off mutant (*tetO-snf21,* 3 hours of inhibition, (Materne et al., 2015)) and control strains on the right panel are presented as an heatmap, where blue represents a gain in nucleosome occupancy and red represents a loss within a region ranging from -750 bp to +750 bp around the TSS at single nucleotide resolution. Rows represent genes and are organized into 5 groups by k-means clustering.

B. Box plot representing the difference in NDR lenght between the *rsc1*□ and *tetO-snf21* strains and the corresponding wt strains. Statistical significance was calculated by a one-sample Wilcoxon test (*pval* < 0.01).

C. Frequency of genes dowregulated (log2 fold change < -0.5, based on (Monahan et al., 2008) in *rsc1Δ* strain sorted by cluster. Cluster 3 includes the *ste11* gene and is enriched (Fisher's exact test p-value < 0.05) for genes whose expression is downregulated in the *rsc1Δ* strain.

APPENDIX II

Yague-Sanz, C., and Hermand, D.

SL-quant: a fast and flexible pipeline to quantify spliced leader trans-splicing events from RNA-seq data.

2018, Gigascience *7*.

In the second part of this thesis, we found that the genes located in position two and over in operons were CDK-12-dependent. As a peculiarity for these genes is to be trans-spliced with the SL2 RNA, we wanted to explore our RNA-seq data in search for such SL trans-splicing events. However, as no tool to quantify trans-splicing events were available, we developed our own, and wanted to share it with the community.

ABSTRACT

The spliceosomal transfer of a short spliced leader (SL) RNA to an independent pre-mRNA molecule is called SL trans-splicing and is widespread in the nematode Caenorhabditis elegans. While RNA-sequencing (RNA-seq) data contain information on such events, properly documented methods to extract them are lacking.

To address this, we developed SL-quant, a fast and flexible pipeline that adapts to paired-end and single-end RNA-seq data and accurately quantifies SL trans-splicing events. It is designed to work downstream of read mapping and uses the reads left unmapped as primary input. Briefly, the SL sequences are identified with high specificity and are trimmed from the input reads, which are then remapped on the reference genome and quantified at the nucleotide position level (SL trans-splice sites) or at the gene level.

SL-quant completes within 10 minutes on a basic desktop computer for typical C. elegans RNA-seq datasets and can be applied to other species as well. Validating the method, the SL trans-splice sites identified display the expected consensus sequence, and the results of the gene-level quantification are predictive of the gene position within operons. We also compared SL-quant to a recently published SL-containing read identification strategy that was found to be more sensitive but less specific than SL-quant. Both methods are implemented as a bash script available under the MIT license [1]. Full instructions for its installation, usage, and adaptation to other organisms are provided.

GigaScience, 7, 2018, 1–7

doi: 10.1093/gigascience/giy084
Advance Access Publication Date: 11 July 2018
Technical Note

TECHNICAL NOTE

SL-quant: a fast and flexible pipeline to quantify spliced leader trans-splicing events from RNA-seq data

Carlo Yague-Sanz * and Damien Hermand

URPhyM-GEMO, The University of Namur (UNamur), 61 rue de Bruxelles, 5000 Namur, Belgium

*Correspondence address. Carlo Yague-Sanz, The University of Namur (UNamur), 61 rue de Bruxelles, 5000 Namur, Belgium. E-mail: carlo.yaguesanz@unamur.be http://orcid.org/0000-0002-9941-9703

ABSTRACT

Background: The spliceosomal transfer of a short spliced leader (SL) RNA to an independent pre-mRNA molecule is called SL trans-splicing and is widespread in the nematode *Caenorhabditis elegans*. While RNA-sequencing (RNA-seq) data contain information on such events, properly documented methods to extract them are lacking. **Findings:** To address this, we developed *SL-quant*, a fast and flexible pipeline that adapts to paired-end and single-end RNA-seq data and accurately quantifies SL trans-splicing events. It is designed to work downstream of read mapping and uses the reads left unmapped as primary input. Briefly, the SL sequences are identified with high specificity and are trimmed from the input reads, which are then remapped on the reference genome and quantified at the nucleotide position level (SL trans-splice sites) or at the gene level. **Conclusions:** *SL-quant* completes within 10 minutes on a basic desktop computer for typical *C. elegans* RNA-seq datasets and can be applied to other species as well. Validating the method, the SL trans-splice sites identified display the expected consensus sequence, and the results of the gene-level quantification are predictive of the gene position within operons. We also compared *SL-quant* to a recently published SL-containing read identification strategy that was found to be more sensitive but less specific than *SL-quant*. Both methods are implemented as a bash script available under the MIT license [1]. Full instructions for its installation, usage, and adaptation to other organisms are provided.

Keywords: NGS; RNA-seq; maturation; trans-splicing; sequence analysis

Background

The capping, splicing, and polyadenylation of eukaryotic pre-mRNAs are well-studied maturation processes that are essential for proper gene expression in eukaryotes [2]. Much less is known about spliced leader (SL) trans-splicing, a process by which a capped small nuclear RNA called spliced leader is spliced onto the 5' end of a pre-mRNA molecule, substituting for canonical capping [3] (Fig. 1A). SL trans-splicing has a patchy phylogenetic distribution ranging from protists [4] to bilaterian metazoans, including nematodes, rotifers [5], and even chordates [6]. It appears not conserved in mammals, although "non-SL" trans-splicing events—when exons from two different RNA transcripts are spliced together—have been detected at low frequency [7]. In contrast, SL trans-splicing is widespread in the *Caenorhabditis elegans* nematode where there are two classes of SL, SL1 and SL2, which trans-splice about 70% of the mRNA transcripts. Strikingly, the SL2 trans-splicing is highly specific for genes in position two and over within operons that range from two to eight genes expressed from a single promoter [8].

While the function of SL trans-splicing begins to be elucidated [9], its regulation remains unclear. To study this question, two main strategies have been proposed to exploit RNA-sequencing (RNA-seq) data in order to quantify SL trans-splicing. The first one involves the mapping of the reads to a complex database containing all the possible trans-spliced gene models [10, 11]. The creation of such a database requires the *in silico*

Received: 18 April 2018; **Revised:** 4 June 2018; **Accepted:** 1 July 2018

© The Author(s) 2018. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/gigascience/article-abstract/7/7/giy084/5052207 by guest on 09 October 2018

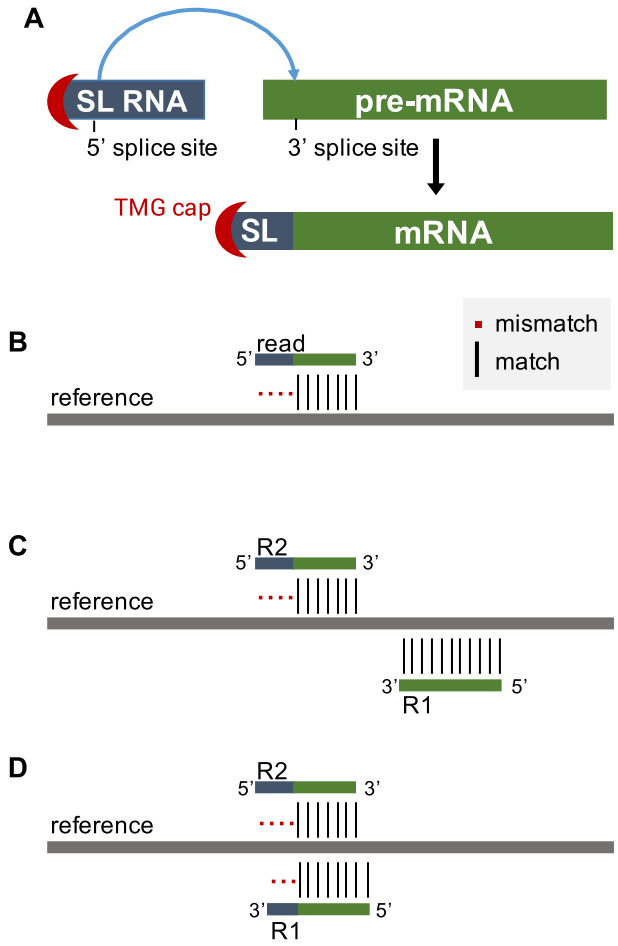


Figure 1: Trans-splicing and RNA-seq. **(A)** The trans-splicing process. Splice leader RNA precursors (SL RNA) are small nuclear RNAs capped with a trimethylguanosine (TMG). The 5'-region of the SL RNA, including the TMG cap, is spliced on the first exon of the pre-mRNAs. **(B)** Reads originating from trans-spliced RNA fragments do not map end-to-end to the reference genome. **(C)** The left-most read (R2) of a read pair does not map end-to-end to the reference. **(D)** Special case when the paired-end reads "dovetail" and both reads do not map end-to-end to the reference due to the SL sequence.

trans-splicing of every SL sequence isoform (12 in *C. elegans*) to all the putative trans-splice sites predicted for a gene. In contrast, the second strategy does not rely on trans-splice site annotation or prediction. Instead, the SL sequences are directly identified in reads partially mapped to the genome or transcriptome [12-14]. However, no implementation of these methods is directly available, which prompted us to develop, test, and optimize *SL-quant*, a ready-to-use pipeline that applies the second strategy to rapidly quantify SL trans-splicing events from RNA-seq data.

Pipeline overview

In order to search for SL sequences in a limited number of reads, only unmapped reads are used as input for *SL-quant*, assuming that reads containing the SL sequence (or the 3' end of it) would not map on the reference genome or transcriptome (Fig. 1B). This implies that a first round of mapping must precede the use of *SL-quant*. It must be performed end-to-end in order to guarantee that reads originating from trans-spliced RNA fragments do not map. In addition to this specification, any bam file containing unmapped reads can be fed into *SL-quant*, making it particularly well suited for subsequent analyses of previously generated data.

In the case paired-end reads are available, only the unmapped reads originating from the left-most ends of the fragments are considered. In addition, we developed an optimized paired-end mode (*-p –paired* option) that further limits the search for SL-containing reads by filtering out the unmapped reads whose mates are also unmapped. This assumes that only the left-most read of a pair originating from a trans-spliced fragment would not map due to the SL sequence while the other one would map (Fig. 1C). This is generally true unless the fragment is so small that the mates significantly overlap with each other (Fig. 1D).

To identify SL trans-splicing events, the input reads are aligned locally to the SL sequences with Basic Local Alignment Search Tool (BLAST) [15]. Reads whose 5' end belongs to a significant alignment (e-value <5%) that covers the 3′ end of the SL sequence (Fig. 2A, left panel) are considered SL-containing reads. Then, the SL-containing reads are trimmed of the SL sequence (based on the length of the BLAST alignment) and mapped back on the *C. elegans* genome with HISAT2 [16]. Finally, the remapped reads are counted at the gene level with *featureCounts* [17] to obtain a quantification of the SL1 and SL2 trans-splicing events per genes.

SL-containing reads identification

We tested *SL-quant* on the single-end modENCODE_4594 [18] dataset (2.5×10^6 unmapped reads) and the paired-end SRR1585277 [19] dataset (1.3×10^6 unmapped left reads) using a desktop computer with basic specifications. Every run was completed within 10 minutes using four threads, with a processing rate of about 10^6 unmapped reads by 5 minutes.

In order to assess the specificity of the BLAST alignments, we reasoned that reads originating from a trans-spliced RNA would align to the 3' end of the SL sequence from their 5' end, while random alignment would start anywhere (Fig. 2A). The fact that 94% of significant alignments were in that specific configuration indicates good specificity (Table 1 and Fig. 2B). In contrast, we obtained less than 0.3% with randomly generated reads. In paired-end mode, fewer alignments were found, but a slightly higher proportion of them (95%) were in proper configuration and considered SL-containing reads. This was expected given the more stringent prefiltering implemented in that mode. When considering only the nonsignificant alignments, we obtained intermediate proportions of proper configuration (15%–20%), suggesting that most, but not all, of those nonsignificant alignments were spurious.

Despite the *C. elegans* SL sequences being 22 nucleotides (nt) long, most alignments cover them on only 10–11 nt (Fig. 2C), with a preference for 10 nt alignment for SL1-containing reads and 11 nt alignments for SL2-containing reads. This could be caused by reverse transcriptase drop-off during the library preparation due to secondary structure and the proximity of the hypermethylated cap at the 5' end of the SL. Moreover, in classic RNA-seq library preparation protocols, the second-strand synthesis is primed by RNA oligonucleotides generated by the digestion of the RNA-DNA duplex obtained after the first strand synthesis. This results in truncated dsDNA fragments that do not preserve the 5' end of the original RNA fragments [20].

Downloaded from https://academic.oup.com/gigascience/article-abstract/7/7/giy084/5052207 by guest on 09 October 2018

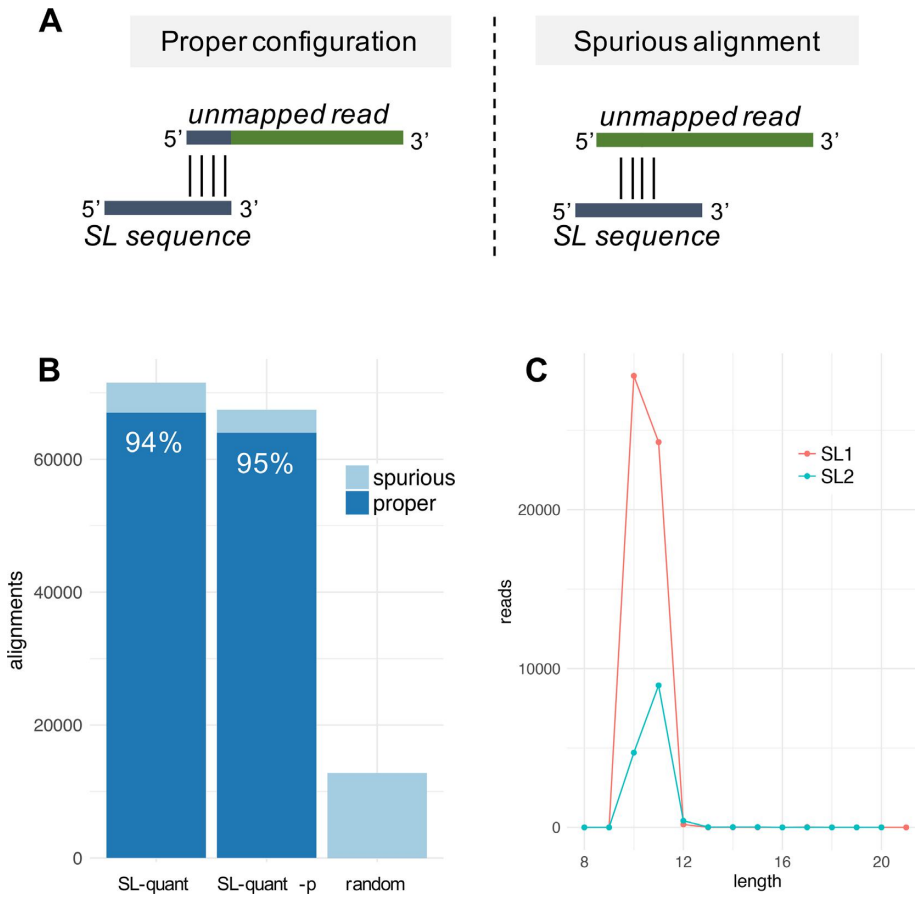


Figure 2: Configuration of the BLAST alignments. **(A)** In *SL-quant*, the BLAST alignments are considered as properly configured if starting from the 5' end of the unmapped read and ending at the 3' end of the SL sequence. **(B)** Proportion of properly configured alignments out of the significant alignment identified by *SL-quant* in single and paired-end (-p) mode on the *SRR1585277* dataset, or on 10^6 random reads in single-end mode. **(C)** Number of properly configured significant alignments found by *SL-quant* on the *SRR1585277* dataset (single-end mode) by alignment length on the SL1 or SL2 sequences.

Table 1: Identification of SL-containing reads by *SL-quant*

				Significant alignments		Nonsignificant alignments	
Dataset	Method	Total reads	Input reads	Total	Properly configured	Total	Properly configured
SRR1585277	SL-quant	40×10^6	1.3×10^6	71, 512	67,021 (94%)	70 211	10,359 (15%)
	SL-quant -p	40×10^6	0.9×10^6	67, 463	64,010 (95%)	47 596	9,849 (21%)
modENCODE_4594	SL-quant	30×10^6	2.5×10^6	168, 351	158,529 (94%)	100 139	20,417 (20%)
random	SL-quant	1×10^6	1.0×10^6	12, 788	36 (0.3%)	43 501	83 (0.2%)

SL-containing reads are defined as reads with significant and properly configured alignment to the SL sequences (sixth column).

SL trans-splice sites identification

While we designed SL-quant with the idea of quantifying SL trans-splicing events by gene, it is also possible to use it to identify the 3' trans-splice sites at single-nucleotide resolution. SL trans-splice sites are known to display the same UUUCAG consensus as *cis*-splice sites [21], which could be verified with our method (Fig. 3A, 3B). Previous work described a significant switch from A to G after consensus sequence (position +1) for the SL1 trans-splice sites compared to SL2 trans-splice sites [21]. At that position, we observed a decreased preference for A for the SL1 trans-splice sites, but no significant enrichment in G. This discrepancy could be due to the fact that we identified (and included in the consensus) about 20 times more SL1 trans-splice sites than previously reported.

As SL trans-splice sites (and splice sites in general) contain an almost invariant AG sequence, we reasoned that non-AG splice sites were potential "spurious" trans-splice sites. In order to assess the performances of our method, we considered identified sites bearing the "AG" consensus as true positives (TPs). Reciprocally, we considered any other sites identified as false positives (FPs), although we cannot completely exclude the existence of nonconsensus splice sites. These reasonable approximations allow us to characterize our method despite not knowing the ground truth. Indicating excellent specificity (ability to exclude FP), 98% of the sites identified by *SL-quant* display the AG con-

Downloaded from https://academic.oup.com/gigascience/article-abstract/7/7/giy084/5052207 by guest on 09 October 2018

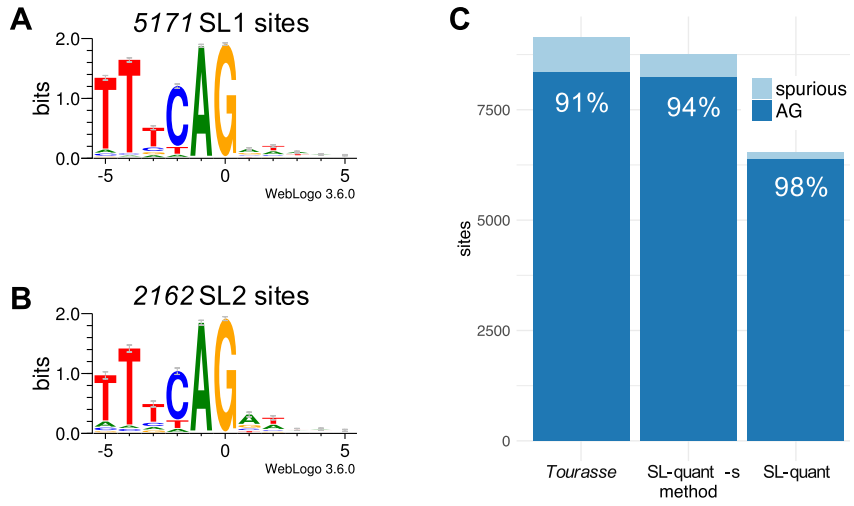


Figure 3: SL-sites consensus sequence. **(A)** Sequence logo of the sequence environment surrounding SL1 or **(B)** SL2 trans-splice sites determined by *SL-quant* on the *SRR1585277* dataset in single-end mode. **(C)** Proportion of AG sequences in SL trans-splice sites identified by *SL-quant* on the *SRR1585277* dataset with the method used in Tourasse et al. 2017 [14] and with *SL-quant* in single-end mode with or without the sensitive option (-s).

Table 2: Performances of *SL-quant* with various parameters.

Dataset	Method	Run time	Mapped SL-containing reads	Trans-splice sites	Site is "AG" consensus (%)
SRR1585277	*SL-quant*	4 minutes 02 seconds	65,126	6,301	6,149 (98)
	SL-quant -p	5 minutes 14 seconds	61,451	6,539	6,402 (98)
	SL-quant -s	2 minutes 45 seconds	120,542	8,770	8,254 (94)
	SL-quant -s -p	6 minutes 58 seconds	114,948	8,436	7,957 (94)
	Tourasse	4 minutes 45 seconds	120,710	8,932	8,260 (92)
modENCODE_4594	*SL-quant*	9 minutes 51 seconds	146,358	8,247	8,081 (98)
	SL-quant -s	3 minutes 10 seconds	258,706	10,735	9,948 (93)
	Tourasse	5 minutes 08 seconds	259,284	11,155	9,953 (89)
random	*SL-quant*	3 minutes 20 seconds	53	52	34 (65)
	SL-quant -s	1m23s	5,757	5,692	5,612 (99 [a])
	Tourasse	2m24s	8,890	8,777	5,612 (64)

[a]The very high proportion of "AG" sites for the random dataset is an artifact caused by the fact that the reads were generated from randomly sampling the genome and that all the *C. elegans* SL sequences end by AG. -p: paired-end mode; -s: sensitive mode.

sensus, regardless of the mode used (single or paired) and the dataset studied (Table 2).

Comparison with a previous method

We also compared our method with a re-implementation of the SL-containing read identification strategy previously reported [14]. Briefly, the unmapped reads whose 5' end align to the SL sequences (or their reverse complement) on at least 5 nt with at most 10% mismatch are considered SL-containing reads. The alignment is realized with *cutadapt* [22] that directly trims the SL sequences from the unmapped reads so they can be remapped to the genome.

Compared to *SL-quant*, this conceptually similar method was faster and identified almost twice the number of SL-containing reads from the real datasets and 150 times the number of SL-containing reads from random reads (Table 2). More splice-sites were identified, but the proportion of spurious (nonconsensus) trans-splice sites increased almost 5-fold (Fig. 3C).

The method developed in [14] has a higher detection power but appears less specific than *SL-quant*. Nevertheless, we consider it an interesting option for applications requiring more sensitivity (ability to detect TP) than specificity. Therefore, we decided to re-implement it within *SL-quant* as an *[-s –sensitive]* option with the following enhancement:

(i) The input reads, if strand specific, are aligned to the SL sequences only (not their reverse complement).
(ii) With paired-end data in single-end mode, only the left-most unmapped reads are considered as input.
(iii) With paired-end data in paired-end mode, only the left-most unmapped reads whose mates are mapped are considered as input.

These modifications significantly improved the specificity of the method (although not to the level of *SL-quant*), with almost no compromise on sensitivity regarding SL trans-splice site detection (Fig. 3C) or SL-containing read identification (Table 2).

Gene-level quantification

Finally, we tested *SL-quant* for its ability to predict gene position within operons as SL2-trans-splicing is the best predictor of transcription initiated upstream of another gene [11] (Fig. 4A). Using the ratio of *SL2/(SL1 + SL2)* from the *SL-quant* output as a

Downloaded from https://academic.oup.com/gigascience/article-abstract/7/7/giy084/5052207 by guest on 09 October 2018

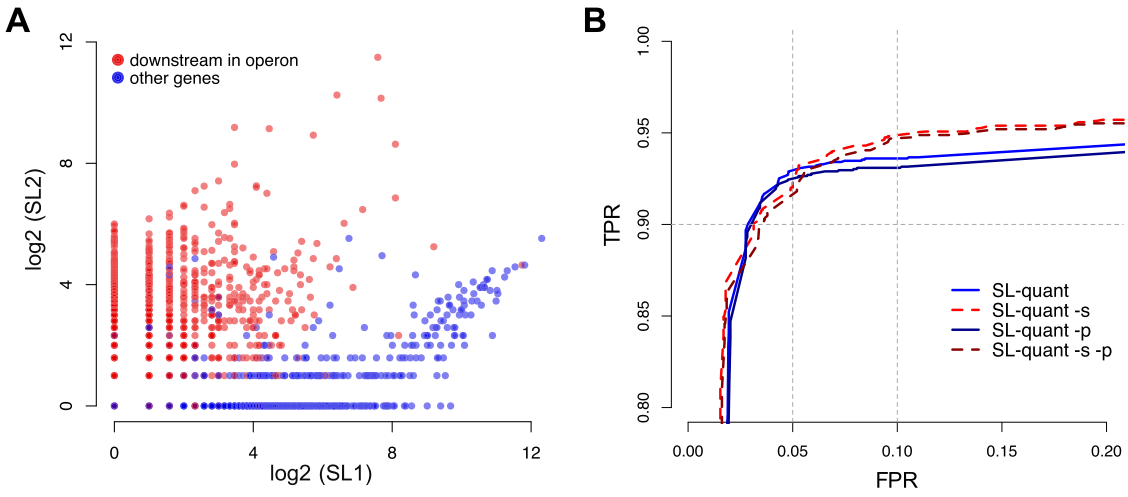


Figure 4: Prediction of gene position in operons. **(A)** Number of SL1 and SL2 trans-splicing events by genes as calculated using *SL-quant*. Genes annotated as downstream in the operons are represented as red dots. **(B)** Receiver operating characteristic curve analysis using the SL2/(SL1 + SL2) ratio as a predictor of downstream position in operons for the 5,521 genes with at least one trans-splicing event detected. The number of SL1 and SL2 trans-splicing events by genes was calculated using *SL-quant* in single or paired (-p) mode, with or without the sensitive (-s) option. TPR: true-positive rate, FPR: false-positive rate.

predictor of gene positions in operons, receiver operating characteristic curve analysis reveals a high TP rate (>90%) at a 5% false discovery rate threshold, regardless of *SL-quant* options (Fig. 4B). However, when tolerating more FPs, *SL-quant* in *sensitive* mode is a superior predictor.

Conclusion

In summary, *SL-quant* is able to rapidly and accurately quantify trans-splicing events from RNA-seq data. It comes as a well-documented and ready-to-use pipeline in which two main options were implemented to fit the type of input data and the intended usage of the quantification (Fig. 5). Importantly, this work provides a way to test and validate SL trans-splicing quantification methods that might serve as a baseline for future development of such methods.

Recently, the hypothesis that the SL trans-splicing mechanism originates from the last eukaryotic common ancestor has been proposed to explain its broad phylogenetic distribution [23]. Given the number of applicable species, the continuously decreasing cost of RNA-seq experiments, and the thinner line between model and nonmodel organisms, it is likely that the SL trans-splicing will be studied in a growing number of species. Therefore, a procedure to adapt *SL-quant* to species other than *C. elegans*, requiring only a few steps, is detailed online. As a proof of concept, we successfully applied *SL-quant* to six additional RNA-seq libraries from five species (Table 3). In the near future, we anticipate that the application of *SL-quant* to various datasets might become instrumental in unveiling trans-splicing regulation in the model organism *C. elegans* and other organisms.

Methods

We ran *SL-quant* with four threads (default) on the modENCODE_4594, modENCODE_4705, modENCODE_4206 [18], SRR2832497 [24], SRR440441, SRR440557 [25], SRR038724 [26], and SRR1585277 [19] poly-A + datasets using a desktop computer with a 2.8-GHz processor and 8 GB random access memory. The *C. elegans, C. briggsae, C. brenneri, and C. remanei* reference genome and annotation (WS262) were downloaded from wormbase [27]. The *T. brucei* reference genome and annotation (Apr_2005 version) were downloaded from Ensembl [28]. The read mapping steps prior to using *SL-quant* and at the end of the pipeline were performed using *HISAT2* [16] (v 2.0.5) with parameters –no-softclip –no-discordant –min-intronlen 20 –max-intronlen 5000. As we noticed adaptor contamination in the modENCODE_4594 dataset, *trimmomatic* [29] (v 0.36) was used to trim them off prior to the mapping. *Samtools* [30] (v 1.5), *picard* [31] (v 2.9), and *bedtools* [32] (v 2.26) were used to convert and/or filter the reads at various stages of the pipeline. BLAST+ (v 2.6) [15] was used to align the reads locally to the relevant SL sequences [33, 34] with parameter -task blastn -word_size 8 max_target_seqs 1. Alternatively, *cutadapt* (v 1.14) [22] was used to directly trim the SL sequences from the reads with parameters -O 5 -m 15 –discard-untrimmed. *FeatureCounts* [17] was used to summarize re-mapped SL-containing reads at the gene level. *Bedtools* [32] was used to summarize mapped SL-containing reads at the genomic position level and to generate random reads by randomly sampling the *C. elegans* genome for 50-nt segments. Sequence logo were made with *weblogo* [35]. Finally, R [36] (v 3.4) was used for analyzing and visualizing the data.

Availability of source code and requirements

Project name: SL-quant
Project home page: https://github.com/cyaguesa/SL-quant
Operating system(s): UNIX-based systems (tested on macOS 10.12.6, macOS 10.11.6, Ubuntu 14.04)
Programming language: Shell, R
Other requirements: The BLAST+ suite (2.6.0 or higher), samtools (1.5 or higher), picard-tools (2.9.0 or higher), featureCounts from the subread package. (1.5.0 or higher), bedtools (2.26.0 or higher), cutadapt (1.14 or higher), hisat2 (2.0.5 or higher). Installation instruction for those requirements is provided online.
License: MIT
RRID:SCR_016205

Downloaded from https://academic.oup.com/gigascience/article-abstract/7/7/giy084/5052207 by guest on 09 October 2018

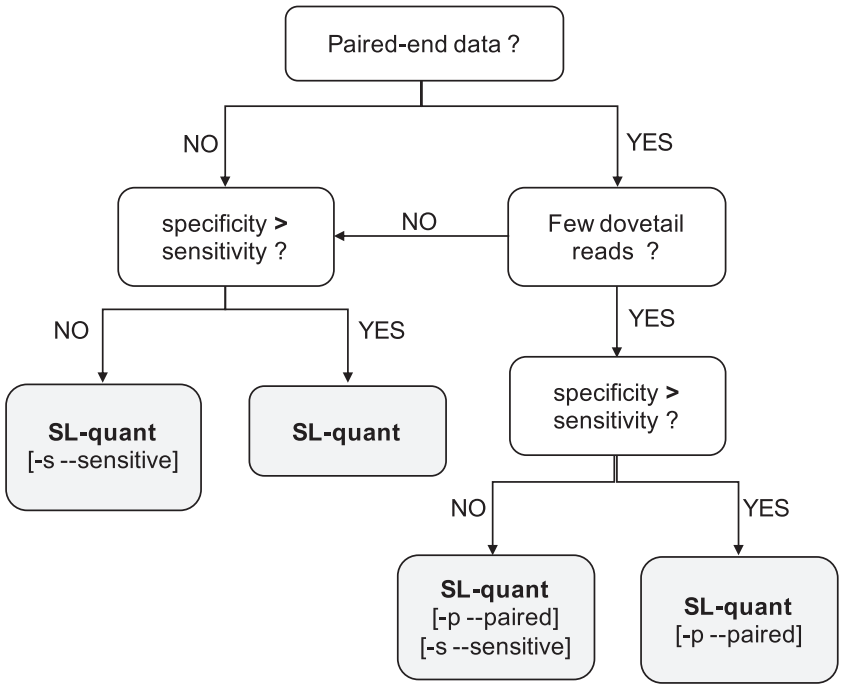


Figure 5: Recommendations on *SL-quant* usage. [-s –sensitive]: it provides increased detection power at the cost of some specificity and it is significantly faster. It is not recommended for applications that are very sensitive to FPs (e.g., trans-splice sites detection) but is an interesting option otherwise (e.g., gene-level quantification of SL trans-splicing events). [-p –paired]: a more stringent prefiltering reduces the number of reads aligned to the SL sequences. It can only be used with paired-end reads. It is not recommended when the average fragment size is small (many "dovetail" reads). It can be used in combination with the [-s –sensitive] option.

Table 3: *SL-quant* can be applied to a wide range of datasets from various species, with varying read length and made with various library preparation protocols.

Organism	Dataset	Read length (nt)	Total reads	Input reads	Mapped SL-containing reads	Trans-splice sites (% AG)
Caenorhabditis elegans	SRR1585277	76	40×10^6	1.3×10^6	120,542	8,770 (94)
	modENCODE_4594	76	30×10^6	2.5×10^6	258,706	10,735 (93)
	SRR2832497 (*)	41	4×10^6	1.8×10^6	16,307	4,882 (87)
Caenorhabditis briggsae	SRR440441	42	11×10^6	5.7×10^6	117,738	8,382 (93)
	SRR440557	42	12×10^6	4.8×10^6	176,205	11,495 (92)
Caenorhabditis brenneri	modENCODE_4705	76	4×10^6	0.4×10^6	74,689	8,891 (97)
Caenorhabditis remanei	modENCODE_4206	76	9×10^6	1.8×10^6	248,335	11,223 (92)
Trypanosoma brucei	SRR038724	35	8×10^6	2.2×10^6	40,320	6,703 (89)

The datasets modENCODE_4594, SRR2832497, and SRR038724 are single end, the others are paired. The asterisk (*) for the SRR2832497 denotes that the second-strand synthesis was made using a ligation-based protocol instead of the classic random priming protocol. All datasets were analyzed with the same *SL-quant* parameters: single-end mode with the -s –sensitive option

Availability of supporting data

The datasets supporting the results presented here are available in the modMine or the European Nucleotide Archive (ebi-ENA) repositories, under the identifiers modENCODE_4594, modENCODE_4705, modENCODE_4206, SRR1585277, SRR2832497, SRR440441, SRR440557, SRR038724. Snapshots of the code and other supporting data are available in the *GigaScience* repository, GigaDB [1].

Additional file

Figure S1. (A) Number of properly oriented significant alignments found by SL-quant on the *SRR2832497* dataset (single-end mode) by alignment length on the SL1 or SL2 sequences. **(B)** Number of properly oriented significant alignments found by with the method used in *Tourasse et al, 2017* on the *SRR1585277* dataset by alignment length on the SL1 or SL2 sequences.

Downloaded from https://academic.oup.com/gigascience/article-abstract/7/7/giy084/5052207 by guest on 09 October 2018

Abbreviations

BLAST: Basic Local Alignment Search Tool; FP: false positive; nt: nucleotide; RNA-seq: RNA sequencing; SL: spliced leader; TP: true positive.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Fond National pour la Recherche Scientifique - Fond pour la formation à la Recherche dans l'Industrie et dans l'Agriculture (FNRS-FRIA).

Author contributions

C.Y. designed, implemented, and tested the pipeline. D.H. and C.Y. wrote the manuscript. D.H. supervised the project.

Acknowledgements

We thank Olivier Finet and Fanélie Bauer for critical reading of the manuscript.

References

1. Yague-Sanz C, Hermand D. Supporting data for "SL-quant: a fast and flexible pipeline to quantify spliced leader trans-splicing events from RNA-seq data." GigaScience Database. 2018. http://dx.doi.org/10.5524/100477.
2. Bentley DL. Coupling mRNA processing with transcription in time and space. Nat Rev Genet 2014;**15**(3):163–75.
3. Blumenthal T. Trans-splicing and operons in *C. elegans*. WormBook 2012:1–11. doi:10.1895/wormbook.1.5.2.
4. Michaeli S. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. Future Microbiol 2011;**6**(4):459–74.
5. Pouchkina-Stantcheva NN, Tunnacliffe A. Spliced leader RNA-mediated trans-splicing in phylum Rotifera. Mol Biol Evol 2005;**22**(6):1482–9.
6. Vandenberghe AE, Meedel TH, Hastings KE. mRNA 5'-leader trans-splicing in the chordates. Genes & Development 2001;**15**(3):294–303.
7. Mangul S, Yang HT, Strauli N, et al. ROP: dumpster diving in RNA-sequencing to find the source of 1 trillion reads across diverse adult human tissues. Genome Biol 2018;**19**(1):36.
8. Blumenthal T, Evans D, Link CD, et al. A global analysis of *Caenorhabditis elegans* operons. Nature 2002;**417**(6891):851–4.
9. Yang YF, Zhang X, Ma X, et al. Trans-splicing enhances translational efficiency in *C. elegans*. Genome Res 2017;**27**(9):1525–35.
10. Hillier LW, Reinke V, Green P, et al. Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. Genome Res 2009;**19**(4):657–66.
11. Allen MA, Hillier LW, Waterston RH, et al. A global analysis of *C. elegans* trans-splicing. Genome Res 2011;**21**(2):255–64.
12. Maxwell CS, Antoshechkin I, Kurhanewicz N, et al. Nutritional control of mRNA isoform expression during developmental arrest and recovery in *C. elegans*. Genome Res 2012;**22**(10):1920–9.
13. Boeck ME, Huynh C, Gevirtzman L, et al. The time-resolved transcriptome of *C. elegans*. Genome Res 2016;**26**(10):1441–50.
14. Tourasse NJ, Millet JRM, Dupuy D. Quantitative RNA-seq meta-analysis of alternative exon usage in *C. elegans*. Genome Res 2017;**27**(12):2120–8.
15. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;**10**:421.
16. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 2015;**12**(4):357–60.
17. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;**30**(7):923–30.
18. Gerstein MB, Lu ZJ, Van Nostrand EL, et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. Science 2010;**330**(6012):1775–87.
19. Kosmaczewski SG, Edwards TJ, Han SM, et al. The RtcB RNA ligase is an essential component of the metazoan unfolded protein response. EMBO Rep 2014;**15**(12):1278–85.
20. Agarwal S, Macfarlan TS, Sartor MA, et al. Sequencing of first-strand cDNA library reveals full-length transcriptomes. Nat Commun 2015;**6**:6002.
21. Graber JH, Salisbury J, Hutchins LN, et al. *C. elegans* sequences that control trans-splicing and operon pre-mRNA processing. RNA 2007;**13**(9):1409–26.
22. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal 2011;**17**(1):10–12.
23. Krchnakova Z, Krajcovic J, Vesteg M. On the possibility of an early evolutionary origin for the spliced leader Trans-Splicing. J Mol Evol 2017;**85**(1–2):37–45.
24. Ni JZ, Kalinava N, Chen E, et al. A transgenerational role of the germline nuclear RNAi pathway in repressing heat stress-induced transcriptional activation in *C. elegans*. Epigenetics Chromatin 2016;**9**:3.
25. Uyar B, Chu JS, Vergara IA, et al. RNA-seq analysis of the *C. briggsae* transcriptome. Genome Res 2012;**22**(8):1567–80.
26. Kolev NG, Franklin JB, Carmi S, et al. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog 2010;**6**(9):e1001090.
27. Stein L, Sternberg P, Durbin R, et al. WormBase: network access to the genome and biology of *Caenorhabditis elegans*. Nucleic Acids Res 2001;**29**(1):82–6.
28. Kersey PJ, Allen JE, Allot A, et al. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. Nucleic Acids Res 2018;**46**(D1):D802–D8.
29. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;**30**(15):2114–20.
30. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;**25**(16):2078–9.
31. Picard: A set of command line tools for manipulating high-troughput sequencing data. http://broadinstitute.github.io/picard. Accessed on 13 Jul 2018.
32. Quinlan AR. BEDTools: The Swiss-Army tool for genome feature analysis. Curr Protoc Bioinformatics 2014;**47**:11.12.1–34.
33. Guiliano DB, Blaxter ML. Operon conservation and the evolution of trans-splicing in the phylum Nematoda. PLos Genet 2006;**2**(11):e198.
34. Bitar M, Boroni M, Macedo AM, et al. The spliced leader trans-splicing mechanism in different organisms: molecular details and possible biological roles. Front Genet 2013;**4**:199.
35. Crooks GE, Hon G, Chandonia JM, et al. WebLogo: a sequence logo generator. Genome Res 2004;**14**(6):1188–90.
36. R Core Team. R: A Language and Environment for Statistical Computing Vienna, Austria: R Foundation for Statistical Computing, 2017.

Downloaded from https://academic.oup.com/gigascience/article-abstract/7/7/giy084/5052207 by guest on 09 October 2018

REFERENCES

Abascal-Palacios, G., Ramsay, E.P., Beuron, F., Morris, E., and Vannini, A. (2018). Structural basis of RNA polymerase III transcription initiation. Nature *553*, 301-306.

Åkerfelt, M., Morimoto, R.I., and Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. Nat Rev Mol Cell Biol *11*, 545--555.

Alaimo, A., Gorojod, R.M., and Kotler, M.L. (2011). The extrinsic and intrinsic apoptotic pathways are involved in manganese toxicity in rat astrocytoma C6 cells. Neurochem Int *59*, 297-308.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). The RNA World and the Origins of Life. Garland Science.

Allen, M.A., Hillier, L.W., Waterston, R.H., and Blumenthal, T. (2011). A global analysis of C. elegans trans-splicing. Genome research *21*, 255-264.

Allison, L.A., Wong, J.K., Fitzpatrick, V.D., Moyle, M., and Ingles, C.J. (1988). The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. Molecular and cellular biology *8*, 321--329.

Anandhakumar, J., Fauquenoy, S., Materne, P., Migeot, V., and Hermand, D. (2013). Regulation of entry into gametogenesis by Ste11: the endless game. Biochem Soc Trans *41*, 1673--1678.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Arimbasseri, A.G., Blewett, N.H., Iben, J.R., Lamichhane, T.N., Cherkasova, V., Hafner, M., and Maraia, R.J. (2015). RNA Polymerase III Output Is Functionally Linked to tRNA Dimethyl-G26 Modification. PLoS genetics *11*, e1005671.

Arimbasseri, A.G., and Maraia, R.J. (2015). Mechanism of Transcription Termination by RNA Polymerase III Utilizes a Non-template Strand Sequence-Specific Signal Element. Molecular cell *58*, 1124--1132.

Arimbasseri, A.G., Rijal, K., and Maraia, R.J. (2014). Comparative overview of RNA polymerase II and III transcription cycles, with focus on RNA polymerase III termination and reinitiation. Transcription *5*, e27639.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Bähler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. Yeast *14*, 943--951.

Baptista, T., Grünberg, S., Minoungou, N., Koster, M.J.E., Timmers, H.T.M., Hahn, S., Devys, D., and Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. Molecular cell *68*, 130--1435.

Barski, A., Chepelev, I., Liko, D., Cuddapah, S., Fleming, A.B., Birch, J., Cui, K., White, R.J., and Zhao, K. (2010). Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. Nature structural & molecular biology *17*, 629-634.

Baskaran, R., Dahmus, M.E., and Wang, J.Y. (1993). Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. PNAS *90*, 11167--11171.

Batta, K., Zhang, Z., Yen, K., Goffman, D.B., and Pugh, B.F. (2011). Genome-wide function of H2B ubiquitylation in promoter and genic regions. Genes & development *25*, 2254--2265.

Bauer, F., Matsuyama, A., Candiracci, J., Dieu, M., Scheliga, J., Wolf, D.A., Yoshida, M., and Hermand, D. (2012). Translational control of cell division by Elongator. Cell Rep *1*, 424--433.

Baugh, L.R. (2013). To grow or not to grow: nutritional control of development during Caenorhabditis elegans L1 arrest. Genetics *194*, 539--555.

Belch, Y., Yang, J., Liu, Y., Malkaram, S.A., Liu, R., Riethoven, J.-J.M., and Ladunga, I. (2010). Weakly positioned nucleosomes enhance the transcriptional competency of chromatin. PloS one *5*, e12984.

Beltchev, B., Yaneva, M., and Staynov, D. (1976). Thermal melting curves of tRNAPhe from yeast lacking different numbers of nucleotides from the 3'-end. Eur J Biochem *64*, 507--510.

Berger, S.L. (2002). Histone modifications in transcriptional regulation. Curr Opin Genet Dev *12*, 142--148.

Betat, H., and Mörl, M. (2015). The CCA-adding enzyme: A central scrutinizer in tRNA quality control. Bioessays *37*, 975--982.

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., Cimermancic, P., Ule, J., and Peterlin, B.M. (2011a). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes & development *25*, 2158-2172.

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., Cimermancic, P., Ule, J., and Peterlin, B.M. (2011b). The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes & development *25*, 2158--2172.

Blumenthal, T., Davis, P., and Garrido-Lecca, A. (2015). Operon and non-operon gene clusters in the C. elegans genome. WormBook *1-20*.

Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M.*, et al.* (2002). A global analysis of Caenorhabditis elegans operons. Nature *417*, 851.

Boguta, M. (2012). Maf1, a general negative regulator of RNA polymerase III in yeast. BBA *1829*, 376--384.

Boguta, M., Czerska, K., and Zoładek, T. (1997). Mutation in a new gene MAF1 affects tRNA suppressor efficiency in Saccharomyces cerevisiae. Gene *185*, 291--296.

Boisnard, S., Lagniel, G., Garmendia-Torres, C., Molin, M., Boy-Marcotte, E., Jacquet, M., Toledano, M.B., Labarre, J., and Chédin, S. (2009). H2O2 activates the nuclear localization of Msn2 and Maf1 through thioredoxins in Saccharomyces cerevisiae. Eukaryot Cell *8*, 1429--1438.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Bowman, E.A., Bowman, C.R., Ahn, J.H., and Kelly, W.G. (2013). Phosphorylation of RNA polymerase II is independent of P-TEFb in the C. elegans germline. Development *140*, 3703-3713.

Buratowski, S. (2003). The CTD code. Nature structural & molecular biology *10*, 679.

Burnol, A.F., Margottin, F., Huet, J., Almouzni, G., Prioleau, M.N., Méchali, M., and Sentenac, A. (1993). TFIIIC relieves repression of U6 snRNA transcription by chromatin. Nature *362*, 475--477.

Byerly, L., Cassada, R.C., and Russell, R.L. (1976). The life cycle of the nematode Caenorhabditis elegans. I. Wild-type growth and reproduction. Dev Biol *51*, 23--33.

Cafferkey, R., McLaughlin, M.M., Young, P.R., Johnson, R.K., and Livi, G.P. (1994). Yeast TOR (DRR) proteins: amino-acid sequence alignment and identification of structural motifs. Gene *141*, 133--136.

Calonge, T.M., Eshaghi, M., Liu, J., Ronai, Z.e., and O'Connell, M.J. (2010). Transformation/Transcription Domain-Associated Protein (TRRAP)-Mediated Regulation of Wee1. Genetics *185*, 81--93.

Canella, D., Bernasconi, D., Gilardi, F., LeMartelot, G., Migliavacca, E., Praz, V., Cousin, P., Delorenzi, M., and Hernandez, N. (2012). A multiplicity of factors contributes to selective RNA polymerase III occupancy of a subset of RNA polymerase III genes in mouse liver. Genome research *22*, 666-680.

Carlsten, J.O.P., Zhu, X., López, M.D., Samuelsson, T., and Gustafsson, C.M. (2016). Loss of the Mediator subunit Med20 affects transcription of tRNA and other non-coding RNA genes in fission yeast. BBA *1859*, 339--347.

Carriere, L., Graziani, S., Alibert, O., Ghavi-Helm, Y., Boussouar, F., Humbertclaude, H., Jounier, S., Aude, J.C., Keime, C., Murvai, J.*, et al.* (2012). Genomic binding of Pol III transcription machinery and relationship with TFIIS transcription factor distribution in mouse embryonic stem cells. Nucleic acids research *40*, 270-283.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.-J., Anderson, S., Yates, J., Washburn, M.P.*, et al.* (2005). Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. Cell *123*, 581--592.

Carter, R., and Drouin, G. (2010). The increase in the number of subunits in eukaryotic RNA polymerase III relative to RNA polymerase II is due to the permanent recruitment of general transcription factors. Mol Biol Evol *27*, 1035--1043.

Castel, S.E., Ren, J., Bhattacharjee, S., Chang, A.Y., Sanchez, M., Valbuena, A., Antequera, F., and Martienssen, R.A. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. Cell *159*, 572-583.

Chatterjee, K., Majumder, S., Wan, Y., Shah, V., Wu, J., Huang, H.-Y., and Hopper, A.K. (2017). Sharing the load: Mex67–Mtr2 cofunctions with Los1 in primary tRNA nuclear export. Genes & development *31*, 2186--2198.

Chatterjee, K., Nostramo, R.T., Wan, Y., and Hopper, A.K. (2018). tRNA dynamics between the nucleus, cytoplasm and mitochondrial surface: Location, location, location. Biochim Biophys Acta, Gene Regul Mech *1861*, 373--386.

Chen, M., and Gartenberg, M.R. (2014). Coordination of tRNA transcription with export at nuclear pore complexes in budding yeast. Genes & development *28*, 959--970.

Chen, Y., Zhang, L., Estarás, C., Choi, S.H., Moreno, L., Karn, J., Moresco, J.J., Yates, J.R., and Jones, K.A. (2014). A gene-specific role for the Ssu72 RNAPII CTD phosphatase in HIV-1 Tat transactivation. Genes & development *28*, 2261--2275.

Chereji, R.V., Ocampo, J., and Clark, D.J. (2017). MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. Molecular cell *65*, 565--577.e563.

Cherkasova, V., Maury, L.L., Bacikova, D., Pridham, K., Bähler, J., and Maraia, R.J. (2011). Altered nuclear tRNA metabolism in La-deleted Schizosaccharomyces pombe is accompanied by a nutritional stress response involving Atf1p and Pcr1p that is suppressible by Xpo-t/Los1p. Mol Biol Cell *23*, 480--491.

Cheung, M.-S., Down, T.A., Latorre, I., and Ahringer, J. (2011). Systematic bias in high-throughput sequencing data and its correction by BEADS. Nucleic acids research *39*, e103.

Cho, E.J., Kobor, M.S., Kim, M., Greenblatt, J., and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. Genes & development *15*, 3319--3329.

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

Ciesla, M., and Boguta, M. (2008). Regulation of RNA polymerase III transcription by Maf1 protein. Acta biochimica Polonica *55*, 215-225.

Cismowski, M.J., Laff, G.M., Solomon, M.J., and Reed, S.I. (1995). KIN28 encodes a C-terminal domain kinase that controls mRNA transcription in Saccharomyces cerevisiae but lacks cyclin-dependent kinase-activating kinase (CAK) activity. Molecular and cellular biology *15*, 2983--2992.

Consortium, C.e.S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012--2018.

Copela, L.A., Fernandez, C.F., Sherrer, R.L., and Wolin, S.L. (2008). Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. RNA *14*, 1214--1227.

Corden, J.L. (1990). Tails of RNA polymerase II. Trends in biochemical sciences *15*, 383-387.

Corden, J.L., Cadena, D.L., Ahearn, J.M., and Dahmus, M.E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. PNAS *82*, 7934--7938.

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

Crick, F.H. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138--163.

Custódio, N., and Carmo-Fonseca, M. (2016). Co-transcriptional splicing and the CTD code. Crit Rev Biochem Mol Biol *51*, 395--411.

David, C.J., Boyne, A.R., Millhouse, S.R., and Manley, J.L. (2011). The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. Genes & development *25*, 972--983.

David, C.J., and Manley, J.L. (2011). The RNA polymerase C-terminal domain: a new role in spliceosome assembly. Transcription *2*, 221--225.

de Almeida, S.F., García-Sacristán, A., Custódio, N., and Carmo-Fonseca, M. (2010). A link between nuclear RNA surveillance, the human exosome and RNA polymerase II transcriptional termination. Nucleic acids research *38*, 8015--8026.

Desai, N., Lee, J., Upadhya, R., Chu, Y., Moir, R.D., and Willis, I.M. (2005). Two steps in Maf1-dependent repression of transcription by RNA polymerase III. The Journal of biological chemistry *280*, 6455--6462.

Dieci, G., Hermann-Le Denmat, S., Lukhtanov, E., Thuriaux, P., Werner, M., and Sentenac, A. (1995). A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. The EMBO journal *14*, 3766-3776.

Dieci, G., Percudani, R., Giuliodori, S., Bottarelli, L., and Ottonello, S. (2000). TFIIIC-independent in vitro transcription of yeast tRNA genes. J Mol Biol *299*, 601--613.

Dittmar, K.A., Goodenbour, J.M., and Pan, T. (2006). Tissue-Specific Differences in Human Transfer RNA Expression. PLoS genetics *2*, e221.

Egel, R. (1973). Commitment to meiosis in fission yeast. Molec Gen Genet *121*, 277--284.

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

Ehara, H., Sekine, S.-i., and Yokoyama, S. (2011). Crystal structure of the C17/25 subcomplex from Schizosaccharomyces pombe RNA polymerase III. Protein Sci *20*, 1558--1565.

Ekumi, K.M., Paculova, H., Lenasi, T., Pospichalova, V., Bösken, C.A., Rybarikova, J., Bryja, V., Geyer, M., Blazek, D., and Barboric, M. (2015). Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. Nucleic acids research *43*, 2575--2589.

Fan, X., Moqtaderi, Z., Jin, Y., Zhang, Y., Liu, X.S., and Struhl, K. (2010). Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. PNAS *107*, 17945--17950.

Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T.G., Nordborg, M., and Lynch, M. (2015). The Spontaneous Mutation Rate in the Fission Yeast Schizosaccharomyces pombe. Genetics *201*, 737--744.

Ferreira Ruiz, M.J., and Umerez, J. (2018). "Dealing with the changeable and blurry edges of living things: a modified version of property-cluster kinds". Euro Jnl Phil Sci *8*, 493--518.

Folco, H.D., Pidoux, A.L., Urano, T., and Allshire, R.C. (2008). Heterochromatin and RNAi Are Required to Establish CENP-A Chromatin at Centromeres. Science *319*, 94--97.

Forsburg, S.L. (2003). Growth and manipulation of S. pombe. Curr Protoc Mol Biol *Chapter*, Unit.

Forsburg, S.L., and Nurse, P. (1991). Cell Cycle Regulation in the Yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe. Annu Rev Cell Biol *7*, 227--256.

Fowler, K.R., Sasaki, M., Milman, N., Keeney, S., and Smith, G.R. (2014). Evolutionarily diverse determinants of meiotic DNA break and recombination landscapes across the genome. Genome research *24*, 1650--1664.

Furuhashi, H., Takasaki, T., Rechtsteiner, A., Li, T., Kimura, H., Checchi, P.M., Strome, S., and Kelly, W.G. (2010). Trans-generational epigenetic regulation of C. elegans primordial germ cells. Epigenetics Chromatin *3*, 15.

Gal, C., Murton, H.E., Subramanian, L., Whale, A.J., Moore, K.M., Paszkiewicz, K., Codlin, S., Bähler, J., Creamer, K.M., Partridge, J.F., *et al.* (2016). Abo1, a conserved bromodomain AAA-ATPase, maintains global nucleosome occupancy and organisation. EMBO Rep *17*, 79--93.

Garrido-Lecca, A., Saldi, T., and Blumenthal, T. (2016). Localization of RNAPII and 3' end formation factor CstF subunits on C. elegans genes and operons. Transcription *7*, 96--110.

Geiger, S.R., Lorenzen, K., Schreieck, A., Hanecker, P., Kostrewa, D., Heck, A.J.R., and Cramer, P. (2010). RNA polymerase I contains a TFIIF-related DNA-binding subcomplex. Molecular cell *39*, 583--594.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., *et al.* (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science *330*, 1775-1787.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., *et al.* (1996). Life with 6000 genes. Science *274*, 546.

Gogakos, T., Brown, M., Garzia, A., Meyer, C., Hafner, M., and Tuschl, T. (2017). Characterizing Expression and Processing of Precursor and Mature Human tRNAs by Hydro-tRNAseq and PAR-CLIP. Cell Rep *20*, 1463--1475.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. PNAS *89*, 5547--5551.

Govind, C.K., Zhang, F., Qiu, H., Hofmeyer, K., and Hinnebusch, A.G. (2007). Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. Molecular cell *25*, 31--42.

Graber, J.H., Salisbury, J., Hutchins, L.N., and Blumenthal, T. (2007). C. elegans sequences that control trans-splicing and operon pre-mRNA processing. RNA *13*, 1409-1426.

Graham, A.C., Kiss, D.L., and Andrulis, E.D. (2009). Core exosome-independent roles for Rrp6 in cell cycle progression. Mol Biol Cell *20*, 2242--2253.

Greenleaf, A.L. (1993). Positive patches and negative noodles: linking RNA processing to transcription? Trends in biochemical sciences *18*, 117--119.

Greig, D., and Leu, J.-Y. (2009). Natural history of budding yeast. Current biology : CB *19*, 886--890.

Gu, B., Eick, D., and Bensaude, O. (2013). CTD serine-2 plays a critical role in splicing and termination factor recruitment to RNA polymerase II in vivo. Nucleic acids research *41*, 1591--1603.

Gudipati, R.K., Xu, Z., Lebreton, A., Séraphin, B., Steinmetz, L.M., Jacquier, A., and Libri, D. (2012). Extensive Degradation of RNA Precursors by the Exosome in Wild-Type Cells. Molecular cell *48*, 409--421.

Guillemette, B., Drogaris, P., Lin, H.-H.S., Armstrong, H., Hiragami-Hamada, K., Imhof, A., Bonneil, E., Thibault, P., Verreault, A., and Festenstein, R.J. (2011). H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. PLoS genetics *7*, e1001354.

Hamada, M., Huang, Y., Lowe, T.M., and Maraia, R.J. (2001). Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. Molecular and cellular biology *21*, 6870--6881.

Hamada, M., Sakulich, A.L., Koduru, S.B., and Maraia, R.J. (2000). Transcription termination by RNA polymerase III in fission yeast. A genetic and biochemically tractable model system. The Journal of biological chemistry *275*, 29076--29081.

Han, L., and Phizicky, E.M. (2018). A rationale for tRNA modification circuits in the anticodon loop. RNA *24*, 1277--1284.

Hanes, S.D. (2014). The Ess1 prolyl isomerase: Traffic cop of the RNA polymerase II transcription. BBA *1839*, 316--333.

Hannon, G.J., Maroney, P.A., and Nilsen, T.W. (1991). U small nuclear ribonucleoprotein requirements for nematode cis- and trans-splicing in vitro. The Journal of biological chemistry *266*, 22792--22795.

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

Heidemann, M., Hintermair, C., Voß, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. BBA *1829*, 55--62.

Helmlinger, D., Marguerat, S., Villén, J., Swaney, D.L., Gygi, S.P., Bähler, J., and Winston, F. (2011). Tra1 has specific regulatory roles, rather than global functions, within the SAGA co-activator complex. The EMBO journal *30*, 2843--2852.

Helmlinger, D., and Tora, L. (2017). Sharing the SAGA. Trends in biochemical sciences *42*, 850--861.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci USA *108*, 18318--18323.

Hennig, B.P., Bendrin, K., Zhou, Y., and Fischer, T. (2012). Chd1 chromatin remodelers maintain nucleosome organization and repress cryptic transcription. EMBO Rep *13*, 997--1003.

Hessle, V., von Euler, A., González de Valdivia, E., and Visa, N. (2012). Rrp6 is recruited to transcribed genes and accompanies the spliced mRNA to the nuclear pore. RNA *18*, 1466--1474.

Hirose, Y., Tacke, R., and Manley, J.L. (1999). Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. Genes & development *13*, 1234--1239.

Ho, C.K., Schwer, B., and Shuman, S. (1998). Genetic, Physical, and Functional Interactions between the Triphosphatase and Guanylyltransferase Components of the Yeast mRNA Capping Apparatus. Molecular and cellular biology *18*, 5189--5198.

Hoffman, C.S., Wood, V., and Fantes, P.A. (2015). An Ancient Yeast for Young Geneticists: A Primer on the Schizosaccharomyces pombe Model System. Genetics *201*, 403--423.

Hoffmann, A., Fallmann, J., Vilardo, E., Mörl, M., Stadler, P.F., and Amman, F. (2018). Accurate mapping of tRNA reads. Bioinformatics *34*, 1116--1124.

Hoffmann, N.A., Jakobi, A.J., Moreno-Morcillo, M., Glatt, S., Kosinski, J., Hagen, W.J.H., Sachse, C., and Müller, C.W. (2015). Molecular structures of unbound and transcribing RNA polymerase III. Nature *528*, 231--236.

Hong, J.C. (2016). Chapter 3 - General Aspects of Plant Transcription Factor Families. Plant Transcription Factors, 35--56.

Hopper, A.K., and Huang, H.-Y. (2015). Quality Control Pathways for Nucleus-Encoded Eukaryotic tRNA Biosynthesis and Subcellular Trafficking. Molecular and cellular biology *35*, 2052.

Hsin, J.-P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing. Science *334*, 683--686.

Hsin, J.-P., Xiang, K., and Manley, J.L. (2014). Function and control of RNA polymerase II C-terminal domain phosphorylation in vertebrate transcription and RNA processing. Molecular and cellular biology *34*, 2488--2498.

Hu, H.-L., Wu, C.-C., Lee, J.-C., and Chen, H.-T. (2015). A Region of Bdp1 Necessary for Transcription Initiation That Is Located within the RNA Polymerase III Active Site Cleft. Molecular and cellular biology *35*, 2831--2840.

Huang, H., Kawamata, T., Horie, T., Tsugawa, H., Nakayama, Y., Ohsumi, Y., and Fukusaki, E. (2015). Bulk RNA degradation by nitrogen starvation-induced autophagy in yeast. The EMBO journal *34*, 154--168.

Huang, Y., and Maraia, R.J. (2001). Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human. Nucleic acids research *29*, 2675--2690.

Hubbard, E.J.A., and Greenstein, D. (2005). Introduction to the germ line (WormBook).

Huisinga, K.L., and Pugh, B.F. (2004). A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. Molecular cell *13*, 573--585.

Hull, M.W., Erickson, J., Johnston, M., and Engelke, D.R. (1994). tRNA genes as transcriptional repressor elements. Molecular and cellular biology *14*, 1266--1277.

Inada, M., Nichols, R.J., Parsa, J.Y., Homer, C.M., Benn, R.A., Hoxie, R.S., Madhani, H.D., Shuman, S., Schwer, B., and Pleiss, J.A. (2016). Phospho-site mutants of the RNA Polymerase II C-terminal domain alter subtelomeric gene expression and chromatin modification state in fission yeast. Nucleic acids research *44*, 9180-9189.

Iwasaki, O., Tanaka, A., Tanizawa, H., Grewal, S.I.S., and Noma, K.-I. (2010). Centromeric localization of dispersed Pol III genes in fission yeast. Mol Biol Cell *21*, 254--265.

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., *et al.* (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074--1080.

Johnson, S.M. (2010). Painting a perspective on the landscape of nucleosome positioning. J Biomol Struct Dyn *27*, 795--802.

Kadaba, S., Krueger, A., Trice, T., Krecic, A.M., Hinnebusch, A.G., and Anderson, J. (2004). Nuclear surveillance and degradation of hypomodified initiator tRNA(Met) in S. cerevisiae. Genes & development *18*, 1227--1240.

Kajitani, T., Kato, H., Chikashige, Y., Tsutsumi, C., Hiraoka, Y., Kimura, H., Ohkawa, Y., Obuse, C., Hermand, D., and Murakami, Y. (2017). Ser7 of RNAPII-CTD facilitates heterochromatin formation by linking ncRNA to RNAi. Proc Natl Acad Sci USA, 201714579.

Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., LeProust, E.M., Hughes, T.R., Lieb, J.D., Widom, J.*, et al.* (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature *458*, 362--366.

Karkusiewicz, I., Turowski, T.W., Graczyk, D., Towpik, J., Dhungel, N., Hopper, A.K., and Boguta, M. (2011). Maf1 protein, repressor of RNA polymerase III, indirectly affects tRNA processing. The Journal of biological chemistry *286*, 39478-39488.

Kassavetis, G.A., Prakash, P., and Shim, E. (2010). The C53/C37 subcomplex of RNA polymerase III lies near the active site and participates in promoter opening. The Journal of biological chemistry *285*, 2695--2706.

Kenneth, N.S., Ramsbottom, B.A., Gomez-Roman, N., Marshall, L., Cole, P.A., and White, R.J. (2007). TRRAP and GCN5 are used by c-Myc to activate RNA polymerase III transcription. PNAS *104*, 14917--14922.

Keogh, M.-C., Kurdistani, S.K., Morris, S.A., Ahn, S.H., Podolny, V., Collins, S.R., Schuldiner, M., Chin, K., Punna, T., Thompson, N.J.*, et al.* (2005). Cotranscriptional Set2 Methylation of Histone H3 Lysine 36 Recruits a Repressive Rpd3 Complex. Cell *123*, 593--605.

Khoo, S.-K., Wu, C.-C., Lin, Y.-C., Lee, J.-C., and Chen, H.-T. (2014). Mapping the protein interaction network for TFIIB-related factor Brf1 in the RNA polymerase III preinitiation complex. Molecular and cellular biology *34*, 551--559.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods *12*, 357--360.

Kim, T., and Buratowski, S. (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. Cell *137*, 259-272.

Kizer, K.O., Phatnani, H.P., Shibata, Y., Hall, H., Greenleaf, A.L., and Strahl, B.D. (2005). A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. Molecular and cellular biology *25*, 3305--3316.

Klass, M., Wolf, N., and Hirsh, D. (1976). Development of the male reproductive system and sexual transformation in the nematode Caenorhabditis elegans. Dev Biol *52*, 1--18.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research *22*, 568--576.

Komarnitsky, P., Cho, E.J., and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. Genes & development *14*, 2452-2460.

Krchnakova, Z., Krajcovic, J., and Vesteg, M. (2017). On the Possibility of an Early Evolutionary Origin for the Spliced Leader Trans-Splicing. J Mol Evol *85*, 37-45.

Krishnamurthy, S., He, X., Reyes-Reyes, M., Moore, C., and Hampsey, M. (2004). Ssu72 Is an RNA polymerase II CTD phosphatase. Molecular cell *14*, 387--394.

Kubik, S., Bruzzone, M.J., Albert, B., and Shore, D. (2017). A Reply to "MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers," by Chereji et al. Molecular cell *65*, 578--580.

Kubik, S., Bruzzone, M.J., Jacquet, P., Falcone, J.-L., Rougemont, J., and Shore, D. (2015). Nucleosome Stability Distinguishes Two Different Promoter Types at All Protein-Coding Genes in Yeast. Molecular cell *60*, 422--434.

Kujirai, T., Ehara, H., Fujino, Y., Shirouzu, M., Sekine, S.-i., and Kurumizaka, H. (2018). Structural basis of the nucleosome transition during RNA polymerase II passage. Science, eaau9904.

Kumar, Y., and Bhargava, P. (2013). A unique nucleosome arrangement, maintained actively by chromatin remodelers facilitates transcription of yeast tRNA genes. BMC genomics *14*, 402.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357--359.

Lemay, J.-F., Larochelle, M., Marguerat, S., Atkinson, S., Bähler, J., and Bachand, F. (2014). The RNA exosome promotes transcription termination of backtracked RNA polymerase II. Nature structural & molecular biology *21*, 919--926.

Lenstra, T.L., Benschop, J.J., Kim, T., Schulze, J.M., Brabers, N.A.C.H., Margaritis, T., van de Pasch, L.A.L., van Heesch, S.A.A.C., Brok, M.O., Koerkamp, M.J.A.G., *et al.* (2011). The Specificity and Topology of Chromatin Interaction Pathways in Yeast. Molecular cell *42*, 536--549.

Li, B., Howe, L., Anderson, S., Yates, J.R., and Workman, J.L. (2003). The Set2 Histone Methyltransferase Functions through the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II. The Journal of biological chemistry *278*, 8897--8903.

Li, F., Cheng, C., Cui, F., de Oliveira, M.V.V., Yu, X., Meng, X., Intorne, A.C., Babilonia, K., Li, M., Li, B., *et al.* (2014). Modulation of RNA polymerase II phosphorylation downstream of pathogen perception orchestrates plant immunity. Cell Host Microbe *16*, 748--758.

Li, J., Moazed, D., and Gygi, S.P. (2002). Association of the Histone Methyltransferase Set2 with RNA Polymerase II Plays a Role in Transcription Elongation. The Journal of biological chemistry *277*, 49383--49388.

Li, J., Wang, Y., Rao, X., Wang, Y., Feng, W., Liang, H., and Liu, Y. (2017). Roles of alternative splicing in modulating transcriptional regulation. BMC Syst Biol *11*, 89.

Li, X., Chatterjee, N., Spirohn, K., Boutros, M., and Bohmann, D. (2016). Cdk12 Is A Gene-Selective RNA Polymerase II Kinase That Regulates a Subset of the Transcriptome, Including Nrf2 Target Genes. Sci Rep *6*, 21455.

Liang, K., Gao, X., Gilmore, J.M., Florens, L., Washburn, M.P., Smith, E., and Shilatifard, A. (2015). Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing. Molecular and cellular biology *35*, 928--938.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Lickwar, C.R., Rao, B., Shabalin, A.A., Nobel, A.B., Strahl, B.D., and Lieb, J.D. (2009). The Set2/Rpd3S Pathway Suppresses Cryptic Transcription without Regard to Gene Length or Transcription Frequency. PloS one *4*, e4886.

Lieberman, J. (2018). Unveiling the RNA World. N Engl J Med *379*, 1278--1280.

Listerman, I., Bledau, A.S., Grishina, I., and Neugebauer, K.M. (2007). Extragenic accumulation of RNA polymerase II enhances transcription by RNA polymerase III. PLoS genetics *3*, e212.

Liu, L., Wang, J., Rosenberg, D., Zhao, H., Lengyel, G., and Nadel, D. (2018). Fermented beverage and food storage in 13,000 y-old stone mortars at Raqefet Cave, Israel: Investigating Natufian ritual feasting. J Archaeolog Sci: Rep *21*, 783--793.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lu, L., Fan, D., Hu, C.-W., Worth, M., Ma, Z.-X., and Jiang, J. (2016). Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II. Biochemistry *55*, 1149--1158.

Luger, K., Rechsteiner, T.J., Flaus, A.J., Waye, M.M., and Richmond, T.J. (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria. J Mol Biol *272*, 301--311.

Lukoszek, R., Mueller-Roeber, B., and Ignatova, Z. (2013). Interplay between polymerase II- and polymerase III-assisted expression of overlapping genes. FEBS Lett *587*, 3692--3695.

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

Machery, E. (2012). Why I stopped worrying about the definition of life... and why you should as well. Synthese *185*, 145--164.

Mahapatra, S., Dewari, P.S., Bhardwaj, A., and Bhargava, P. (2011). Yeast H2A.Z, FACT complex and RSC regulate transcription of tRNA gene through differential dynamics of flanking nucleosomes. Nucleic acids research *39*, 4023-4034.

Malone, J.H., and Oliver, B. (2011). Microarrays, deep sequencing and the true measure of the transcriptome. BMC Biol *9*, 34.

Marshall, L., and White, R.J. (2012). Non-coding RNA production by RNA polymerase III is implicated in cancer. Nat Rev Cancer *12*, 732.

Mason, A.G., Garza, R.M., McCormick, M.A., Patel, B., Kennedy, B.K., Pillus, L., and La Spada, A.R. (2017). The replicative lifespan-extending deletion of SGF73 results in altered ribosomal gene expression in yeast. Aging Cell *16*, 785--796.

Mata, J., and Wise, J.A. (2017). 4-Thiouridine Labeling to Analyze mRNA Turnover in Schizosaccharomyces pombe. Cold Spring Harb Protoc *2017*, 5.

Materne, P., Anandhakumar, J., Migeot, V., Soriano, I., Yague-Sanz, C., Hidalgo, E., Mignion, C., Quintales, L., Antequera, F., and Hermand, D. (2015). Promoter nucleosome dynamics regulated by signalling through the CTD code. eLife *4*.

Materne, P., Vazquez, E., Sanchez, M., Yague-Sanz, C., Anandhakumar, J., Migeot, V., Antequera, F., and Hermand, D. (2016). Histone H2B ubiquitylation represses gametogenesis by opposing RSC-dependent chromatin remodeling at the ste11 master regulator locus. eLife *5*.

Matsuo, Y., Asakawa, K., Toda, T., and Katayama, S. (2006). A rapid method for protein extraction from fission yeast. Biosci Biotechnol, Biochem *70*, 1992--1994.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome research *18*, 1073--1083.

Maxwell, C.S., Antoshechkin, I., Kurhanewicz, N., Belsky, J.A., and Baugh, L.R. (2012). Nutritional control of mRNA isoform expression during developmental arrest and recovery in C. elegans. Genome research *22*, 1920-1929.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J., and Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. Nature structural & molecular biology *17*, 1272-1278.

Mbonye, U.R., Gokulrangan, G., Datt, M., Dobrowolski, C., Cooper, M., Chance, M.R., and Karn, J. (2013). Phosphorylation of CDK9 at Ser175 enhances HIV transcription and is a marker of activated P-TEFb in CD4(+) T lymphocytes. PLoS Pathog *9*, e1003338.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature *385*, 357--361.

McDaniel, S.L., Hepperla, A.J., Huang, J., Dronamraju, R., Adams, A.T., Kulkarni, V.G., Davis, I.J., and Strahl, B.D. (2017). H3K36 Methylation Regulates Nutrient Stress Response in Saccharomyces cerevisiae by Enforcing Transcriptional Fidelity. Cell Reports *19*, 2371--2382.

McGraw-Hill (2012). Encyclopedia of Science and Technology.

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

Migeot, V., and Hermand, D. (2018). Chromatin Immunoprecipitation-Polymerase Chain Reaction (ChIP-PCR) Detects Methylation, Acetylation, and Ubiquitylation in S. pombe. Methods Mol Biol *1721*, 25--34.

Mizumoto, K., and Kaziro, Y. (1987). Messenger RNA capping enzymes from eukaryotic cells. Prog Nucleic Acid Res Mol Biol *34*, Review.

Moqtaderi, Z., Wang, J., Raha, D., White, R.J., Snyder, M., Weng, Z., and Struhl, K. (2010). Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. Nature structural & molecular biology *17*, 635-640.

Moreno, S., Klar, A., and Nurse, P. (1991). Molecular genetic analysis of fission yeast Schizosaccharomyces pombe. Methods Enzymol *194*, Research.

Morris, D.P., and Greenleaf, A.L. (2000). The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. The Journal of biological chemistry *275*, 39935--39943.

Mortimer, R.K. (2000). Evolution and Variation of the Yeast (Saccharomyces) Genome. Genome research *10*, 403--409.

Mosley, A.L., Pattenden, S.G., Carey, M., Venkatesh, S., Gilmore, J.M., Florens, L., Workman, J.L., and Washburn, M.P. (2009). Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation. Molecular cell *34*, 168--178.

Nagarajavel, V., Iben, J.R., Howard, B.H., Maraia, R.J., and Clark, D.J. (2013). Global 'bootprinting' reveals the elastic architecture of the yeast TFIIIB–TFIIIC transcription complex in vivo. Nucleic acids research *41*, 8135--8143.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2002). Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. Genes & development *16*, 806--819.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Molecular cell *11*, 709--719.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526--540.

Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., and Carmo-Fonseca, M. (2018). RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. Molecular cell *72*, 369-379 e364.

Nonet, M., Sweetser, D., and Young, R.A. (1987). Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell *50*, 909--915.

Nurse, P. (2017). A Journey in Science: Cell-Cycle Control. Mol Med *23*, 112--119.

Oesterreich, F.C., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell *165*, 372--381.

Oler, A.J., Alla, R.K., Roberts, D.N., Wong, A., Hollenhorst, P.C., Chandler, K.J., Cassiday, P.A., Nelson, C.A., Hagedorn, C.H., Graves, B.J., *et al.* (2010). Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. Nature structural & molecular biology *17*, 620-628.

Orioli, A., Praz, V., Lhôte, P., and Hernandez, N. (2016). Human MAF1 targets and represses active RNA polymerase III genes by preventing recruitment rather than inducing long-term transcriptional arrest. Genome research *26*, 624--635.

Otsubo, Y., and Yamamoto, M. (2012). Signaling pathways for fission yeast sexual differentiation at a glance. J Cell Sci *125*, 2789--2793.

Pan, L., Xie, W., Li, K.-L., Yang, Z., Xu, J., Zhang, W., Liu, L.-P., Ren, X., He, Z., Wu, J., *et al.* (2015). Heterochromatin remodeling by CDK12 contributes to learning in Drosophila. PNAS *112*, 13988--13993.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet *40*, 1413--1415.

Pan, T. (2018). Modifications and functional genomics of human transfer RNA. Cell Res *28*, 395--404.

Parnell, T.J., Huff, J.T., and Cairns, B.R. (2008). RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. The EMBO journal *27*, 100-110.

Parnell, T.J., Schlichter, A., Wilson, B.G., and Cairns, B.R. (2015). The chromatin remodelers RSC and ISW1 display functional and chromatin-based promoter antagonism. eLife *4*, e06073.

Pointner, J., Persson, J., Prasad, P., Norman-Axelsson, U., Strålfors, A., Khorosjutina, O., Krietenstein, N., Svensson, J.P., Ekwall, K., and Korber, P. (2012). CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in S. pombe. The EMBO journal *31*, 4388--4403.

Portz, B., Lu, F., Gibbs, E.B., Mayfield, J.E., Mehaffey, M.R., Zhang, Y.J., Brodbelt, J.S., Showalter, S.A., and Gilmour, D.S. (2017). Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. Nat Commun *8*, 15231.

Pradhan, S.K., Xue, Y., and Carey, M.F. (2015). Fragile Nucleosomes Influence Pol II Promoter Function. Molecular cell *60*, 342--343.

Pratt-Hyatt, M., Pai, D.A., Haeusler, R.A., Wozniak, G.G., Good, P.D., Miller, E.L., McLeod, I.X., Yates, J.R., Hopper, A.K., and Engelke, D.R. (2013). Mod5 protein binds to tRNA gene complexes and affects local transcriptional silencing. Proc Natl Acad Sci USA *110*, E3081--E3089.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. Cell *108*, 501--512.

Pugh, B.F., and Venters, B.J. (2016). Genomic Organization of Human Transcription Initiation Complexes. PloS one *11*, e0149339.

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

Quan, T.K., and Hartzog, G.A. (2010). Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of Saccharomyces cerevisiae Spt4-Spt5 in transcription. Genetics *184*, 321--334.

R Core Team (2017). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing).

Raha, D., Wang, Z., Moqtaderi, Z., Wu, L., Zhong, G., Gerstein, M., Struhl, K., and Snyder, M. (2010). Close association of RNA polymerase II and many transcription factors with Pol III genes. Proceedings of the National Academy of Sciences of the United States of America *107*, 3639-3644.

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

Roberts, D.N., Stewart, A.J., Huff, J.T., and Cairns, B.R. (2003). The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. Proceedings of the National Academy of Sciences of the United States of America *100*, 14695-14700.

Roberts, D.N., Wilson, B., Huff, J.T., Stewart, A.J., and Cairns, B.R. (2006). Dephosphorylation and genome-wide association of Maf1 with Pol III-transcribed genes during repression. Molecular cell *22*, 633-644.

Rodríguez-Navarro, S. (2009). Insights into SAGA function during gene expression. EMBO Rep *10*, 843--850.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. Nature *224*, 234-237.

Roux, A.E., Langhans, K., Huynh, W., and Kenyon, C. (2016). Reversible Age-Related Phenotypes Induced during Larval Quiescence in C. elegans. Cell Metab *23*, 1113--1126.

Salat-Canela, C., Paulo, E., Sánchez-Mir, L., Carmona, M., Ayté, J., Oliva, B., and Hidalgo, E. (2017). Deciphering the role of the signal- and Sty1 kinase-dependent phosphorylation of the stress-responsive transcription factor Atf1 on gene activation. The Journal of biological chemistry *292*, 13635--13644.

Sansó, M., Vargas-Pérez, I., Quintales, L., Antequera, F., Ayté, J., and Hidalgo, E. (2011). Gcn5 facilitates Pol II progression, rather than recruitment to nucleosome-depleted stress promoters, in Schizosaccharomyces pombe. Nucleic acids research *39*, 6369--6379.

Schlake, T., and Gutz, H. (1993). Mating configurations in Schizosaccharomyces pombe strains of different geographical origins. Curr Genet *23*, 108--114.

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

Schmitt, M.E., Brown, T.A., and Trumpower, B.L. (1990). A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. Nucleic acids research *18*, 3091--3092.

Schones, D.E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., and Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. Cell *132*, 887--898.

Schröder, S., Herker, E., Itzen, F., He, D., Thomas, S., Gilchrist, D.A., Kaehlcke, K., Cho, S., Pollard, K.S., Capra, J.A., *et al.* (2013). Acetylation of RNA polymerase II regulates growth-factor-induced gene transcription in mammalian cells. Molecular cell *52*, 314--324.

Schwer, B., Bitton, D.A., Sanchez, A.M., Bähler, J., and Shuman, S. (2014). Individual letters of the RNA polymerase II CTD code govern distinct gene expression programs in fission yeast. PNAS *111*, 4185--4190.

Schwer, B., and Shuman, S. (2011). Deciphering the RNA polymerase II CTD code in fission yeast. Molecular cell *43*, 311--318.

Shim, E.Y., Walker, A.K., Shi, Y., and Blackwell, T.K. (2002). CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo. Genes & development *16*, 2135-2146.

Shim, Y.S., Choi, Y., Kang, K., Cho, K., Oh, S., Lee, J., Grewal, S.I.S., and Lee, D. (2012). Hrp3 controls nucleosome positioning to suppress non-coding transcription in eu- and heterochromatin. The EMBO journal *31*, 4375--4387.

Shkreta, L., and Chabot, B. (2015). The RNA Splicing Response to DNA Damage. Biomolecules *5*, 2935--2977.

Shukla, A., and Bhargava, P. (2018). Regulation of tRNA gene transcription by the chromatin structure and nucleosome dynamics. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms *1861*, 295--309.

Sipiczki, M. (2000). Where does fission yeast sit on the tree of life? Genome Biol *1*, reviews1011.1011--reviews1011.1014.

Skowronek, E., Grzechnik, P., Späth, B., Marchfelder, A., and Kufel, J. (2014). tRNA 3′ processing in yeast involves tRNase Z, Rex1, and Rrp6. RNA *20*, 115--130.

Soutourina, J., Bordas-Le Floch, V., Gendrel, G., Flores, A., Ducrot, C., Dumay-Odelot, H., Soularue, P., Navarro, F., Cairns, B.R., Lefebvre, O., *et al.* (2006). Rsc4 connects the chromatin remodeler RSC to RNA polymerases. Molecular and cellular biology *26*, 4920--4933.

Stevens, A. (1960). Incorporation of the adenine ribonucleotide into RNA by cell fractions from E. coli B. Biochem Biophys Res Commun *3*, 92--96.

Stiller, J.W., and Cook, M.S. (2004). Functional Unit of the RNA Polymerase II C-Terminal Domain Lies within Heptapeptide Pairs. Eukaryotic Cell *3*, 735--740.

Suh, H., Ficarro, S.B., Kang, U.-B., Chun, Y., Marto, J.A., and Buratowski, S. (2016). Direct analysis of phosphorylation sites on the Rpb1 C-terminal domain of RNA polymerase II. Molecular cell *61*, 297--304.

Suh, H., Hazelbaker, D.Z., Soares, L.M., and Buratowski, S. (2013). The C-terminal domain of Rpb1 functions on other RNA polymerase II subunits. Molecular cell *51*, 850--858.

Sukegawa, Y., Yamashita, A., and Yamamoto, M. (2011). The fission yeast stress-responsive MAPK pathway promotes meiosis via the phosphorylation of Pol II CTD in response to environmental and feedback cues. PLoS genetics *7*, e1002387.

Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Molecular cell *52*, 52--62.

Takahara, T., and Maeda, T. (2012). TORC1 of fission yeast is rapamycin-sensitive. Genes Cells *17*, 698--708.

Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. Genome research *18*, 1084--1091.

Van Horn, D.J., Yoo, C.J., Xue, D., Shi, H., and Wolin, S.L. (1997). The La protein in Schizosaccharomyces pombe: a conserved yet dispensable phosphoprotein that functions in tRNA maturation. RNA *3*, 1434--1443.

van Steensel, B. (2011). Chromatin: constructing the big picture. The EMBO journal *30*, 1885--1895.

Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Molecular cell *45*, 439--446.

Vannini, A., Ringel, R., Kusser, A.G., Berninghausen, O., Kassavetis, G.A., and Cramer, P. (2010). Molecular basis of RNA polymerase III transcription repression by Maf1. Cell *143*, 59--70.

Varshney, D., Vavrova-Anderson, J., Oler, A.J., Cairns, B.R., and White, R.J. (2015). Selective repression of SINE transcription by RNA polymerase III. Mob Genet Elements *5*, 86--91.

Venkatesh, S., Li, H., Gogol, M.M., and Workman, J.L. (2016). Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. Nat Commun *7*, 13610.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The Sequence of the Human Genome. Science *291*, 1304--1351.

Vera, D.L., Madzima, T.F., Labonne, J.D., Alam, M.P., Hoffman, G.G., Girimurugan, S.B., Zhang, J., McGinnis, K.M., Dennis, J.H., and Bass, H.W. (2014). Differential Nuclease Sensitivity Profiling of Chromatin Reveals Biochemical Footprints Coupled to Gene Expression and Functional DNA Elements in Maize. Plant Cell *26*, 3883--3893.

Volpe, T., and Martienssen, R.A. (2011). RNA Interference and Heterochromatin Assembly. Cold Spring Harbor Perspect Biol *3*, a003731.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833--1837.

Volschenk, H., van Vuuren, H.J.J., and Viljoen-Bloom, M. (2003). Malo-ethanolic fermentation in Saccharomyces and Schizosaccharomyces. Curr Genet *43*, 379--391.

Vorländer, M.K., Khatter, H., Wetzel, R., Hagen, W.J.H., and Müller, C.W. (2018). Molecular mechanism of promoter opening by RNA polymerase III. Nature *553*, 295--300.

Voss, K., Forné, I., Descostes, N., Hintermair, C., Schüller, R., Maqbool, M.A., Heidemann, M., Flatley, A., Imhof, A., Gut, M.*, et al.* (2015). Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II. Transcription *6*, 91--101.

Wagner, P., Grimaldi, M., and Jenkins, J.R. (1993). Putative dehydrogenase tms1 suppresses growth arrest induced by a p53 tumour mutant in fission yeast. Eur J Biochem *217*, 731--736.

Wan, Y., and Hopper, A.K. (2018). From powerhouse to processing plant: conserved roles of mitochondrial outer membrane proteins in tRNA splicing. Genes & development *32*, 1309--1314.

Wang, X., Wang, X., Li, L., and Wang, D. (2010). Lifespan extension in Caenorhabditis elegans by DMSO is dependent on sir-2.1 and daf-16. Biochem Biophys Res Commun *400*, 613-618.

Warner, J.R. (1999). The economics of ribosome biosynthesis in yeast. Trends in biochemical sciences *24*, 437--440.

Weake, V.M., and Workman, J.L. (2012). SAGA function in tissue-specific gene expression. Trends Cell Biol *22*, 177--184.

West, S., Gromak, N., and Proudfoot, N.J. (2004). Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature *432*, 522--525.

White, R.J. (2011). Transcription by RNA polymerase III: more complex than we thought. Nature reviews Genetics *12*, 459-463.

Wilkinson, M.G., Samuels, M., Takeda, T., Toone, W.M., Shieh, J.C., Toda, T., Millar, J.B., and Jones, N. (1996). The Atf1 transcription factor is a target for the Sty1 stress-activated MAP kinase pathway in fission yeast. Genes & development *10*, 2289--2301.

Wilusz, J.E. (2015). Removing roadblocks to deep sequencing of modified RNAs. Nat Methods *12*, 821.

Wolkow, C.A.H., D.H. (2016). Handbook of C. elegans Dauer Anatomy. WormAtlas.

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S.*, et al.* (2002). The genome sequence of Schizosaccharomyces pombe. Nature *415*, 871.

Wood, V., Harris, M.A., McDowall, M.D., Rutherford, K., Vaughan, B.W., Staines, D.M., Aslett, M., Lock, A., Bahler, J., Kersey, P.J.*, et al.* (2012). PomBase: a comprehensive online resource for fission yeast. Nucleic acids research *40*, D695-699.

Workman, J.L. (2006). Nucleosome displacement in transcription. Genes & development *20*, 2009--2017.

Wu, J., Bao, A., Chatterjee, K., Wan, Y., and Hopper, A.K. (2015). Genome-wide screen uncovers novel pathways for tRNA processing and nuclear-cytoplasmic dynamics. Genes & development *29*, 2633--2644.

Wyce, A., Xiao, T., Whelan, K.A., Kosman, C., Walter, W., Eick, D., Hughes, T.R., Krogan, N.J., Strahl, B.D., and Berger, S.L. (2007). H2B ubiquitylation acts as a barrier to Ctk1 nucleosomal recruitment prior to removal by Ubp8 within a SAGA-related complex. Molecular cell *27*, 275--288.

Xiao, S., Scott, F., Fierke, C.A., and Engelke, D.R. (2001). Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes. Annu Rev Biochem *71*, 165--189.

Xiao, T., Hall, H., Kizer, K.O., Shibata, Y., Hall, M.C., Borchers, C.H., and Strahl, B.D. (2003). Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes & development *17*, 654--663.

Xie, C., and Tammi, M.T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinf *10*, 80.

Yague-Sanz, C., and Hermand, D. (2018). SL-quant: a fast and flexible pipeline to quantify spliced leader trans-splicing events from RNA-seq data. Gigascience *7*.

Yague-Sanz, C., Vazquez, E., Sanchez, M., Antequera, F., and Hermand, D. (2017). A conserved role of the RSC chromatin remodeler in the establishment of nucleosome-depleted regions. Curr Genet *63*, 187-193.

Yang, Y.F., Zhang, X., Ma, X., Zhao, T., Sun, Q., Huan, Q., Wu, S., Du, Z., and Qian, W. (2017). Trans-splicing enhances translational efficiency in C. elegans. Genome research *27*, 1525-1535.

Yeganeh, M., Praz, V., Cousin, P., and Hernandez, N. (2017). Transcriptional interference by RNA polymerase III affects expression of the Polr3e gene. Genes & development *31*, 413--421.

Yoo, C.J., and Wolin, S.L. (1997). The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. Cell *89*, 393--402.

Zaros, C., and Thuriaux, P. (2005). Rpc25, a conserved RNA polymerase III subunit, is critical for transcription initiation. Molecular microbiology *55*, 104--114.

Zaslaver, A., Baugh, L.R., and Sternberg, P.W. (2011). Metazoan operons accelerate recovery from growth-arrested states. Cell *145*, 981--992.

Zehring, W.A., Lee, J.M., Weeks, J.R., Jokerst, R.S., and Greenleaf, A.L. (1988). The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro. PNAS *85*, 3698--3702.

Zhang, C., Lopez, M.S., Dar, A.C., Ladow, E., Finkbeiner, S., Yun, C.H., Eck, M.J., and Shokat, K.M. (2013). Structure-guided inhibitor design expands the scope of analog-sensitive kinase technology. ACS Chem Biol *8*, 1931-1938.

Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M., and Pan, T. (2015). Efficient and quantitative high-throughput tRNA sequencing. Nat Methods *12*, 835.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M.*, et al.* (2018). The interaction landscape between transcription factors and the nucleosome. Nature, 1.