

Université de Montréal

**Les patrons d'expression de gènes : ont-ils évolué avec la
complexité des organismes?**

par

Sandra-Rima Imrazène

Département de Biochimie

Faculté de Médecine

Mémoire présenté à la Faculté de Médecine
en vue de l'obtention du grade de Maître ès Sciences
en Biochimie
option Génétique Moléculaire

Décembre 2013

© Sandra-Rima Imrazène, 2013

Université de Montréal
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé

Les patrons d'expression de gènes : ont-ils évolué avec la complexité des organismes?

Présenté par :

Sandra-Rima Imrazène

évalué par un jury composé des personnes suivantes :

Gerardo Ferbeyre, président-rapporteur

Daniel Zenklusen, directeur de recherche

Pascal Chartrand, membre du jury

Résumé

La régulation de la transcription est l'un des processus cellulaires des plus fondamentaux et constitue la première étape menant à l'expression protéique. Son altération a des effets sur l'homéostasie cellulaire et est associée au développement de maladies telles que le cancer. Il est donc crucial de comprendre les règles fondamentales de la fonction cellulaire afin de mieux cibler les traitements pour les maladies.

La transcription d'un gène peut se produire selon l'un des deux modes fondamentaux de transcription : en continu ou en *burst*. Le premier est décrit comme un processus aléatoire et stochastique qui suit une distribution de Poisson. À chaque initiation de la transcription, indépendante de la précédente, un seul transcrit est produit. L'expression en *burst* se produit lorsque le promoteur est activé pour une courte période de temps pendant laquelle plusieurs transcrits naissants sont produits. Apportant la plus grande variabilité au sein d'une population isogénique, il est représenté par une distribution bimodale, où une sous-population n'exprime pas le gène en question, alors que le reste de la population l'exprime fortement.

Les gènes des eucaryotes inférieurs sont pour la plupart exprimés de manière continue, alors que les gènes des eucaryotes supérieurs le sont plutôt en *burst*. Le but de ce projet est d'étudier comment l'expression des gènes a évolué et si la transcription aléatoire, ou de Poisson, est une propriété des eucaryotes inférieurs et si ces patrons ont changé avec la complexité des organismes et des génomes. Par la technique de smFISH, nous avons étudié de manière systématique quatre gènes évolutivement conservés (*mdn1*⁺, *PRP8/spp42*⁺, *pol1*⁺ et *cdc13*⁺) qui sont continuellement transcrits dans la levure *S. cerevisiae*. Nous avons observé que le mode d'expression est gène-et-organisme spécifique puisque *prp8* est exprimé de manière continue dans la levure *S. pombe*, alors que les autres gènes seraient plutôt exprimés en légers *burst*.

Mots-clés: transcription, expression génique, patron d'expression, continu, *burst*

Abstract

Regulating transcription is one of the most fundamental cellular processes and the first step of a long cascade of processes leading to protein expression. Altering transcriptional output often has major effects on cellular homeostasis and is associated with many disease phenotypes, such as cancer. Understanding the fundamental rules governing transcription regulation is therefore instrumental in understanding cellular function as well as in finding disease treatments.

Transcription of a gene can occur through two fundamental different modes: “continuous” or “bursting”. Continuous transcription is defined as stochastic process where a promoter is always in its “on” state and each initiation event is independent of the previous. Bursting transcription occurs when a promoter is activated for a short time and multiple mRNAs are produced during the “on” state, followed by long periods of transcription inactivity. It leads to greater variability in an isogenic population, as expression is often bimodal as a sub-population does not express a given gene.

Bursting expression is frequently observed in higher eukaryotes, while a continuous pattern seems to be common in lower eukaryotes. The goal of this project is to study how gene expression patterns evolved. We investigate whether Poisson-like transcription is a property of lower eukaryotes and whether transcription patterns have changed when organisms and genomes evolved into more complex systems. Using smFISH, we have systematically determined expression patterns of four evolutionarily conserved genes; *mdn1*⁺, *PRP8/spp42*⁺, *pol1*⁺ and *cdc13*⁺, previously shown to be continuously expressed in the yeast *S. cerevisiae*. Expression was studied in the yeast *S. pombe*, as an example for another lower eukaryote as well as in human cell-lines. We observe that expression patterns are organism-and-gene specific suggesting that expression patterns have evolved to fulfill gene specific functions.

Keywords : gene expression, expression patterns, transcription, continuous, bursting

Table des matières

A. Introduction.....	12
I. Expression des gènes.....	12
Initiation de la transcription	12
Modèle d'assemblage séquentiel du PIC	13
Modèle de l'holoenzyme de l'ARN Polymérase II	13
Élongation et terminaison	17
Variabilité dans les populations isogéniques	18
Corrélation entre les variations transcriptionnelles <i>versus</i> variations traductionnelles	20
Patrons d'expression de gènes	21
i. Continuuel	22
ii. <i>Bursting</i>	23
II. Facteurs influençant l'expression des gènes.....	25
Contexte génomique	25
Pause de l'ARN Polymérase II	26
Séquence et architecture du promoteur	29
<i>Gene-looping</i>	31
III. Évolution de la complexité des génomes.....	32
IV. Patrons d'expression de gènes dans l'évolution.....	34
Patron continuuel chez les organismes eucaryotes inférieurs	34
Patron <i>bursting</i> chez les organismes eucaryotes supérieurs	34
V. Étude des patrons d'expression des gènes par smFISH.....	36
VI. Rationnel.....	39
 B. Materials & Methods.....	 40
C. Results.....	45
D. Discussion.....	74
E. Conclusion et perspectives.....	91

Liste des tableaux

Table 1. Facteur Fano calculé pour chaque stades du cycle cellulaire pour les transcrits naissants et cytoplasmiques de *mdn1*⁺ et le résultat du test de chi-carré d'une distribution de Poisson dans la souche de *S. pombe* 972 h-.

Table 2. Facteur Fano calculé pour chaque stades du cycle cellulaire pour les transcrits naissants et cytoplasmiques de *PRP8/spp42*⁺ et le résultat du test de chi-carré d'une distribution de Poisson dans la souche de *S. pombe* 972 h-.

Table 3. Facteur Fano calculé pour chaque stades du cycle cellulaire pour les transcrits naissants et cytoplasmiques de *pol1*⁺ et *cdc13*⁺, ainsi que le résultat du test de chi-carré d'une distribution de Poisson dans la souche de *S. pombe* 972 h-.

Tableau 4. Tableau comparatif des différents paramètres calculés théoriques et observés pour chacun des gènes étudiés en G2 dans la levure *S. pombe* tels que la longueur, le temps de transcription des gènes, le taux théorique de transcription des gènes par minute et par heure et les moyennes des transcrits naissants observée et le facteur Fano de chaque gène en G2.

Liste des figures

Figure 1 Représentation schématique des deux modèles de formation du complexe de pré-initiation
A. Représentation du modèle et de l'ordre d'assemblage séquentiel du complexe chez les humains ; TFIID est d'abord recruté à la boîte TATA puis, TFIIA et TFIIB sont recrutés afin de stabiliser l'interaction de TFIID à la boîte TATA. L'ARN Pol II avec TFIIF sont recrutés près du site d'initiation. Une fois ces composantes assemblées, TFIIE est recruté, suivi de TFIIH. **B.** Représentation du modèle du recrutement de l'holoenzyme au site promoteur d'un gène. TFIID en complexe avec TFIIA est recruté à la boîte TATA suivi de l'holoenzyme composé de l'ARN Pol II, TFIIB, TFIIF, TFIIE et TFIIH.

Figure 2. Schéma simplifié du phénomène d'amplification de la transcription d'un gène. Des protéines activatrices lient les enhanceurs et font le pont avec la machinerie basale de la transcription grâce à des co-activateurs.

Figure 3. Patrons d'expression de gènes **A.** (Gauche) Représentation des événements d'initiation de la transcription par rapport au temps pour un gène exprimé par le mode continu. En vert sont représentés les transcrits issus d'un événement d'initiation (rectangle bleu). (Droite) Distribution de la fréquence du nombre de transcrits naissants pour un gène transcrit continuellement correspondant aussi à une distribution de Poisson. **B.** (Gauche) Représentation des événements d'initiation de la transcription par rapport au temps pour un gène exprimé par le mode *bursting*. En vert sont représentés les transcrits issus d'un événement d'initiation (rectangle bleu). (Droite) Distribution de la fréquence du nombre de transcrits naissants pour un gène transcrit *burst* correspondant aussi à une distribution bi-modale.

Figure 4. Représentation de la pause de l'ARN Polymérase II après l'initiation de la transcription. De manière générale, les facteurs NELF (Negative Elongation Factor) et DSIF (DRB Sensitive Inducing Factor) sont recrutés et s'associent à l'enzyme, l'empêchant de poursuivre son activité. Une série de phosphorylations et de déphosphorylations par d'autres facteurs permettent la reprise de l'activité de l'ARN Polymérase II.

Figure 5. Représentation schématique du phénomène de ré-initiation de la transcription par *gene looping*. Une boucle est formée entre la région promotrice et l'extrémité 3' d'un gène permettant à l'ARN Polymérase II de réinitier la transcription.

Figure 6. **A.** Représentation schématique des sondes fluorescentes utilisées dans la technique de smFISH hybridées à un ARNm spécifique. **B.** Image superposée obtenue par smFISH de cellules de *S. cerevisiae* hybridées par des sondes ciblant l'ARNm de *mdn1*⁺ en rouge, le noyau visible grâce au DAPI et la forme des cellules par lumière transmise. Les cercles jaunes entourent un ARNm unique de *mdn1*⁺ ainsi qu'un site de transcription où l'accumulation de la fluorescence des sondes hybridées aux transcrits naissants. **C.** Représentation de trois ARNm de longueurs différentes ainsi que la probabilité d'en observer des transcrits naissants et l'intensité relative supposée être observée selon le mode d'expression employé.

Figure 7. Graphic representation of each gene's length and structure, introns are represented in dark grey and coding regions in light green. Left: smFISH merged image of nuclear signal in blue (DAPI) and each gene in red (Cy3.5) for *mdn1*⁺ (A), *PRP8/spp42*⁺ (B), *poll*⁺ (C) and *cdc13*⁺ (D). Right: cells shown on left panel under transmitted light.

Figure 8. Cells segmented by Cell Profiler (A-right) using autofluorescence in GFP channel (A-left) in an epifluorescence microscope. Nuclei segmented by Cell Profiler (B-right) using DAPI signal (B-left). Spots are detected by Localize using 2D Gaussian fit (C-left). ARNQuant, a MATLAB based script (developed by Emmanuel Noutahi) merging of segmented cells, segmented nuclei (C-right).

Figure 9. Left: smFISH experiment merged image of *mdn1*⁺ transcripts (Cy3.5) and nuclei (DAPI) in *S.*

pombe. A. Frequency distributions of nascent (A) transcripts per allele and cytoplasmic transcripts (B) with superposed fitting of a Poisson distribution. The results were quantified from smFISH performed on *S. pombe* 972 h- strain from three independent experiments (n=775 cells, 1552 Alleles).

Figure 10. Left: smFISH merged image of nuclei (DAPI) and *mdn1*⁺ transcripts (Cy3.5). Right: Number of transcription sites of *mdn1*⁺ gene in *S. pombe* 972 h- strain from three independent experiments (n=775 cells).

Figure 11. Left: smFISH merged images of nuclei (DAPI) and *mdn1*⁺ (Cy3.5) in *S. pombe* 972 h- cells. Right: Cells from left panel under transmitted light pointed out at different cell-cycle stages as categorized in this project. *S. pombe* 972 h- strain cells spend around 70% of their lifespan in G2 phase, 10% in M phase and 20% in G1 and S phase.

Figure 12. A. Frequency distribution of nascent transcripts in all alleles separately (n=1070 alleles) of cells in G2 phase only with superposed Poisson distribution. Data quantified from three independent smFISH experiments. B. Frequency distribution of cytoplasmic transcripts of *mdn1*⁺ in *S. pombe* 972 h- strain in G2 stage (n=535 cells) with superposed Poisson distribution. C. Result of nascent transcripts of *mdn1*⁺ distribution data tested with MATLAB function **alldistfit** for fitting discrete distribution. D. Frequency distribution of nascent transcripts of cells in G2 phase only without inactive alleles and Poisson distribution fit superposed.

Figure 13. Left panels: smFISH merged images of nuclei (DAPI) and *mdn1*⁺ (Cy3.5) in *S. pombe* 972 h- cells with, circled in yellow, cells in M phase (top) and G1/S phase (bottom). Frequency distributions of nascent transcripts per allele of *mdn1*⁺ in *S. pombe* 972 h- during M phase (A) and G1/S phase (C) with a fit of a Poisson distribution superposed. Frequency distributions of cytoplasmic transcripts of *mdn1*⁺ in *S. pombe* 972 h- strain in M stage (n=114 cells) (B) and G1/S phase (n=100 cells) (D) with a superposed fit for Poisson distribution. All the results shown are quantified from three independent experiments.

Figure 14. Left: smFISH image of merged nuclei (DAPI) and *mdn1*⁺ (Cy3.5) in *S. pombe* 972 h- cells. Distributions of nascent (A) transcripts and cytoplasmic (C) transcripts of *mdn1*⁺ in *S. pombe* and superposition of a Poisson distribution. Binning and superposition respectively of nascent (B) and cytoplasmic (D) transcripts of *mdn1*⁺ in G2 phase (blue), M phase (grey) and G1/S phase (red).

Figure 15. Left: smFISH experiment merged image of *PRP8/spp42*⁺ transcripts (Cy3.5) and nuclei (DAPI) in *S. pombe*. A. Frequency distributions of nascent (A) transcripts per allele and cytoplasmic transcripts (B) with superposed fitting of a Poisson distribution. The results were quantified from smFISH performed on *S. pombe* 972 h- strain from three independent experiments (n=524 cells).

Figure 16. Left: smFISH merged image of *PRP8/spp42*⁺ transcripts (Cy3.5) and nuclei (DAPI) in *S. pombe* 972 h- strain. Right: Number of transcription sites of *PRP8/spp42*⁺ gene in *S. pombe* 972 h- strain from three independent experiments (n=524 cells).

Figure 17. Left: merged smFISH with *PRP8/spp42*⁺ mRNA (Cy 3.5) and nuclei (DAPI). A cell in G2 phase is circled in yellow. A. Frequency distribution of nascent transcripts per allele of *PRP8/spp42*⁺ in *S. pombe* 972 h- strain in G2 phase and a superposed fitting of a Poisson distribution on the nascent transcripts per allele distribution in G2 (p=0.8205, $\chi^2_{stat}=1.5347$) B. Frequency distribution of cytoplasmic transcripts of *PRP8/spp42*⁺ in *S. pombe* 972 h- strain in G2 stage (n= 392 cells) with superposed fitting of a Poisson distribution.

Figure 18. Left: merged smFISH with *PRP8/spp42*⁺ mRNA (Cy 3.5) and nuclei (DAPI) with yellow circled cell in M phase (top) and G1/S phase (bottom). Frequency distributions of nascent transcripts per allele of *PRP8/spp42*⁺ in *S. pombe* 972 h- strain in M phase (A) and in G1/S phase (C) and a superposed fitting of a

Poisson distribution on the nascent transcripts for M phase ($p=0.1960$ $\chi^2_{\text{stat}}=3.2591$) and G1/S phase ($p=0.2947$ $\chi^2_{\text{stat}}=2.4433$) Frequency distributions of cytoplasmic transcripts of *PRP8/spp42*⁺ in *S. pombe* 972 h- strain in M phase (B) (n= 82 cells) and G1/S phase (D) (n=50 cells) with superposed fittings of a Poisson distribution.

Figure 19. Average nascent transcripts for *mdn1*⁺ (A) and *PRP8/spp42*⁺ (B) in *S. pombe* during M phase at different d/e coefficients from three independent experiments. For *mdn1*⁺ and *PRP8/spp42*⁺ respectively, 2% and 5% of cells were found between 0 and 0.2, 13% and 10% between 0.2 and 0.4, 52% and 56% between 0.4 and 0.6, and 33% and 29% between 0.6 and 1.

Figure 20. Left: smFISH image of merged nuclei (DAPI) and *PRP8/spp42*⁺ (Cy3.5) in *S. pombe* 972 h- cells. Distributions of nascent (A) transcripts and cytoplasmic (C) transcripts of *PRP8/spp42*⁺ in *S. pombe* and superposition of a Poisson distribution. Binning and superposition respectively of nascent (B) and cytoplasmic (D) transcripts of *PRP8/spp42*⁺ in G2 phase (blue), M phase (grey) and G1/S phase (red).

Figure 21. Left panels: smFISH merged image of nuclear signal (DAPI) and *poll*⁺ (top) or *cdc13*⁺ mRNAs (Cy3.5) with respective gene length's and structure representation. Frequency distributions of nascent *poll*⁺ (A) and *cdc13*⁺ (C) transcripts per allele in *S. pombe* 972 h- and fitting of a Poisson distribution for each. Frequency of cytoplasmic transcripts of *poll*⁺ (B) (n= 669 cells) and *cdc13*⁺ (D) (n= 656 cells) and fitting of a Poisson distribution for each. The results were quantified from smFISH performed on *S. pombe* 972 h- strain from three independent experiments.

Figure 22. Left panels: smFISH merged image of nuclei (DAPI) and *poll*⁺ or *cdc13*⁺ transcripts in *S. pombe* cells. Cells found at different cell-cycle stages are circled in yellow for each gene. Frequency of nascent transcripts of *poll*⁺ at phase G2 (A), M (C), and G1/S (E); *cdc13*⁺ transcripts at phase G2 (G), M (I), and G1/S (K) with superposed fitting of Poisson distribution. Frequency of cytoplasmic transcripts of *poll*⁺ at phase G2 (B) (n=498 cells), M (D) (n=92 cells), and G1/S (F) (n=84 cells); *cdc13*⁺ at phase G2 (G) (n=472 cells), M (I) (n=104 cells), and G1/S (K) (n=148 cells) with superposed fitting of Poisson distribution. All the results shown are quantified from three independent experiments.

Figure 23. Left panels: smFISH merged image of nuclei (DAPI) and *poll*⁺ or *cdc13*⁺ transcripts in *S. pombe* cells. Distributions of nascent transcripts of *poll*⁺ (A) and *cdc13*⁺ (E) and cytoplasmic transcripts of *poll*⁺ (C) and *cdc13*⁺ (G) in *S. pombe* and superpositions of a Poisson distribution. Binning and superposition respectively of *poll*⁺ (B) and *cdc13*⁺ (F) nascent as well as *poll*⁺ (D) and *cdc13*⁺ (H) cytoplasmic transcripts in G2 phase (blue), M phase (grey) and G1/S phase (red).

Figure 24. (A) Representation of MDN1 gene with exons (dark red rectangles) and introns and disposition of 5' probe set (red dotted line) and 3' probe set (green dotted line). (B-D) smFISH experiments on Human cell-lines for *mdn1*⁺ gene and nuclear signal by DAPI. (B Left to right) 5' exons in Cy3, 3' exons in Cy3.5 and merged signals for HEK293T cells. (C Left to right) 5' exons in Cy3, 3' exons in Cy3.5 and merged signals for U2OS cells. (D Left to right) 5' exons in Cy3, 3' exons in Cy3.5 and merged signals for IMR90 cells.

Figure 25. 3D quantification of transcripts in Human cell-lines. A-B. Data of detected 5' exonic and 3' exonic spots, obtained in HEK293T cells, from FISH-Quant were imported into Imaris. The images are a superposition of localized spots in 3D and original Cy3 signal of 5' exonic (A) or Cy3.5 3' exonic (B) MDN1 transcripts and DAPI C. Original DAPI signal and colocalized 5' and 3' spots are then considered, by ARNQuant in Imaris developed by Emmanuel Noutahi, as a single and complete transcript. D. Original DAPI, Cy3 for 5' exonic and Cy3.5 for 3' exonic MDN1 transcripts merged with segmented cells and nuclei by Imaris.

Figure 26. Distributions des fréquences d'ARNm par cellules dans *S. cerevisiae* (bleu) provenant de et dans

S. pombe (rouge) obtenue dans le présent projet.

Figure 27. Modèles d'initiation de la transcription avec l'activation du promoteur (ligne bleue) et l'initiation de la transcription est représentée par les barres verticales verte. **A.** Patron d'expression continuuel où le promoteur est actif de manière continue et où la transcription est initiée de manière aléatoire suivant une distribution de Poisson. **B.** Patron d'expression en *burst*, où le temps d'inactivité du promoteur est grand et où le nombre de transcrits naissants par période d'activité est élevé correspondant à une distribution bimodale. **C.** Modèle de *burst* léger où le promoteur possède des périodes d'inactivités et où, par initiation, plus d'un transcrit est produit. Ce modèle ne mène ni à une distribution de Poisson, ni à une distribution bimodale.

Liste des abréviations

<u>Abréviation</u>	<u>Description</u>
ARN	Acide ribonucléique
ARNm	ARN messenger
ARN Pol II	ARN Polymérase II
ADN	Acide désoxyribonucléique
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
<i>S. pombe</i>	<i>Schizosaccharomyces pombe</i>
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
<i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
DAPI	4',6-diamidino-2-phenylindole
p-TEFB	Facteur d'élongation de la transcription positif

*À Vassiliy Iossifovitch Moroz,
ta rigueur scientifique, ton intégrité et ton
éthique de vie guident mes pas*

Remerciements

J'aimerais d'abord remercier Daniel Zenklusen de m'avoir accueillie au sein de son laboratoire et de m'avoir donné la chance de m'épanouir scientifiquement pendant les deux années qu'a duré ma maîtrise. J'ai eu l'opportunité de travailler sur un projet qui m'a passionnée avec des outils à la fine pointe de la technologie, le tout, dans un laboratoire et un département stimulants.

De plus, j'aimerais remercier Emmanuel Noutahi sans qui, je n'aurais jamais pu avancer aussi rapidement dans ce projet. Grâce à son expertise en bio-informatique, nous avons pu mettre sur pieds des outils qui faciliteront les analyses de nos données.

Enfin, et non les moindres, j'aimerais remercier ceux qui m'ont appuyée dans mes années en Biochimie et surtout pendant ma maîtrise : Olga Moroz, Mohand Imrazene, Linda Imrazene, Kader Benghenam et Andrés F. Gil Blanco. Je vous remercie de tout cœur de votre compréhension et votre soutien constants, qui m'ont permis d'accomplir une étape importante de ma vie. Tous vos gestes, des plus petits aux plus grands, ont été plus qu'appréciés.

A. Introduction

I. L'expression des gènes

Initiation de la transcription

Les cellules eucaryotes, contrairement aux cellules procaryotes, possèdent un noyau dans lequel l'information génétique est entreposée dans une molécule d'ADN. Celle-ci est compacté en unités appelées nucléosomes grâce aux histones, protéines sur lesquelles la molécule d'ADN s'enroule. L'ensemble condensé de ces unités constitue la chromatine.

Important centre de commande de la cellule, le noyau est l'organelle dans laquelle les premières étapes de l'expression des gènes, contenus dans la chromatine, se produisent. Dans tout organisme eucaryote, elle suit le dogme central de la biologie moléculaire, c'est-à-dire que l'information contenue dans l'ADN double-brin est transcrite en une version miroir sous forme d'ARN messenger simple-brin pour être ensuite traduite en sa version protéique finale. Cependant, afin que cela se produise, il faut d'abord que la chromatine soit décompactée afin que la machinerie de transcription ait accès à la région dite promotrice du gène. Tous les gènes contiennent une région promotrice indiquant à la machinerie transcriptionnelle l'endroit précis où commencer la transcription. Pour ce faire, des protéines dites remodeleuses de la chromatine permettent de déplacer les histones, désencombrant ainsi l'ADN afin d'entreprendre les étapes d'initiation de la transcription [1-3]

Ce processus est généralement initié par des facteurs de transcription qui sont recrutés de manière séquence-spécifique au promoteur. Ils recrutent à leur tour des facteurs qui remodelent la chromatine et la machinerie générale de la transcription qui initie la production d'un ARNm naissant. Pour le recrutement de la machinerie générale de la transcription au promoteur et l'initiation de la transcription, deux modèles sont proposés : soit par le modèle d'assemblage séquentiel du complexe de pré-initiation (PIC) ou par le modèle de l'holoenzyme de l'ARN polymérase II.

Modèle d'assemblage séquentiel du PIC

La réaction de transcription est accomplie par une enzyme qui a pour fonction d'ajouter des nucléotides à une chaîne d'ARNm naissante, l'ARN Polymérase II. Contrairement aux organismes plus simples, tels que les bactériophages, l'ARN Polymérase II constitue un large complexe macromoléculaire essentiel pour permettre à la transcription d'avoir lieu. Pour que celle-ci puisse être entamée à la région promotrice d'un gène, des facteurs de transcription généraux (TFIIx) sont nécessaires. Ils ont été identifiés et catalogués en sous-complexes ayant des fonctions et un ordre d'assemblage différents.

Plusieurs laboratoires ont démontré que différentes composantes du complexe s'assemblent de manière séquentielle. Ce modèle a été le mieux décrit, *in vitro*, par Buratowski *et al.* en 1989, à l'aide de la technique d'électrophorèse sur gel natif et par la technique d'empreinte à la DNase I [4]. Le facteur de transcription IID (TFIID) reconnaît la boîte TATA sur le promoteur du gène et permet le recrutement consécutif du TFIIA et du TFIIIB qui stabilisent l'ARN Polymérase II et le TFIIIF. Une fois un complexe stable formé, les facteurs TFIIIE et TFIIH sont recrutés (Figure 1A). TFIIIE facilite les activités enzymatiques ATPase et hélicase de TFIIH permettant la création d'une tension superhéliçoidale négative dans l'ADN. Cette étape crée un accès encore plus prononcé et efficace de l'ADN pour l'ARN Polymérase II qui peut enfin débiter la transcription.

Modèle de l'holoenzyme de l'ARN Polymérase II

Le second modèle de production d'un PIC apte à initier la transcription, et le plus accepté actuellement, est celui de l'holoenzyme. Un holoenzyme est un complexe enzymatique actif composé de l'enzyme et de sa ou ses coenzyme(s). Ainsi, ce modèle a été proposé par plusieurs laboratoires qui ont purifié dans le nucléoplasme un complexe d'holoenzyme contenant l'ARN Pol II [5, 6], avec ou sans facteurs de transcription généraux, avec des protéines impliquées dans le remodelage de la chromatine, la réparation de l'ADN et dans le traitement de l'ARN. [2, 7-14] Cet holoenzyme est donc préassemblé et se lie au promoteur comme un méga-complexe permettant l'initiation de la transcription (Figure 1B).

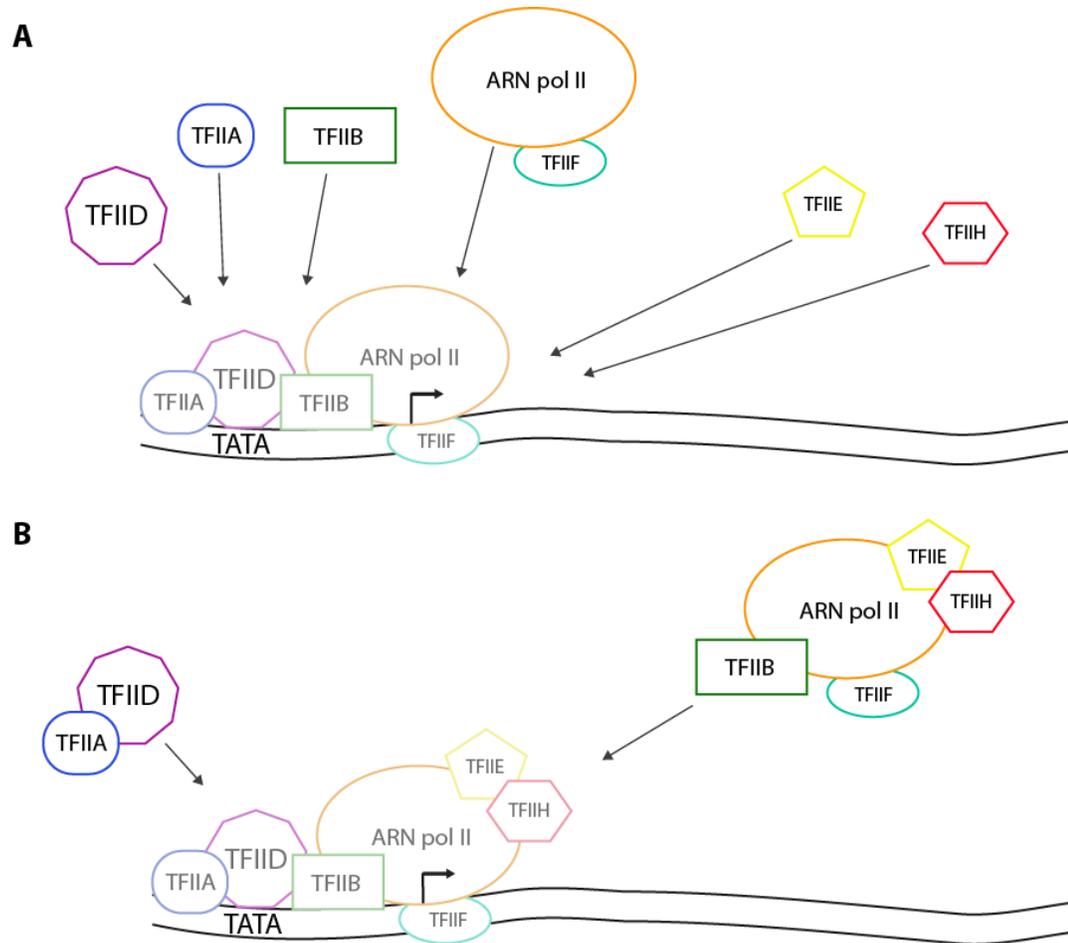


Figure 1 Représentation schématique des deux modèles de formation du complexe de pré-initiation
A. Représentation du modèle et de l'ordre d'assemblage séquentiel du complexe chez les humains (d'après [15] et [16]) TFIID est d'abord recruté à la boîte TATA puis, TFIIA et TFIIB sont recrutés afin de stabiliser l'interaction de TFIID à la boîte TATA. L'ARN Pol II avec TFIIF sont recrutés près du site d'initiation. Une fois ces composantes assemblées, TFIIE est recruté, suivi de TFIIH. **B.** Représentation du modèle du recrutement de l'holoenzyme au site promoteur d'un gène. (Adapté de [16]) TFIID en complexe avec TFIIA est recruté à la boîte TATA suivi de l'holoenzyme composé de l'ARN Pol II, TFIIB, TFIIF, TFIIE et TFIIH.

Il existe, dans les cellules eucaryotes, des éléments permettant d'accroître l'expression d'un gène. Par exemple, les activateurs de transcription sont des protéines se liant, pour la plupart, à des séquences d'ADN spécifiques aux gènes qu'ils activent, à proximité ou directement au promoteur. Ils ont aussi des domaines permettant des liaisons protéines-protéines soit avec des co-activateurs qui font le lien avec les facteurs généraux de la transcription, ou directement à ces derniers. Les activateurs sont souvent recrutés pour des gènes qui doivent être activés suite à un signal spécifique perçu par la cellule (Figure 2). Des répresseurs agissent de manière similaire aux activateurs tout en réprimant l'expression d'un gène plutôt que de l'activer.

Enfin, certains éléments à même l'ADN permettant d'accentuer l'expression d'un gène appelés « enhancers », parfois appelés amplificateurs, existent dans les génomes d'organismes eucaryotes. Ces séquences sont reconnues spécifiquement par des protéines activatrices les liant et permettent à leur tour de lier des protéines régulatrices se situant dans la région promotrice du gène dont l'expression soit être augmentée [17].

Dans la levure *Schizosaccharomyces pombe*, ces « enhancers » sont situés assez près du promoteur, soit autour de 40 paires de bases en amont. Par contre, dans les cellules humaines, certains de ces « enhancers » peuvent être à plus de 1000 kb du promoteur d'un gène à amplifier [18-20].

Ainsi, en plus de la transcription basale d'un gène en question, les cellules eucaryotes sont dotées de séquences d'ADN et de protéines permettant d'augmenter l'expression de certains gènes [2, 6, 16, 18-21].

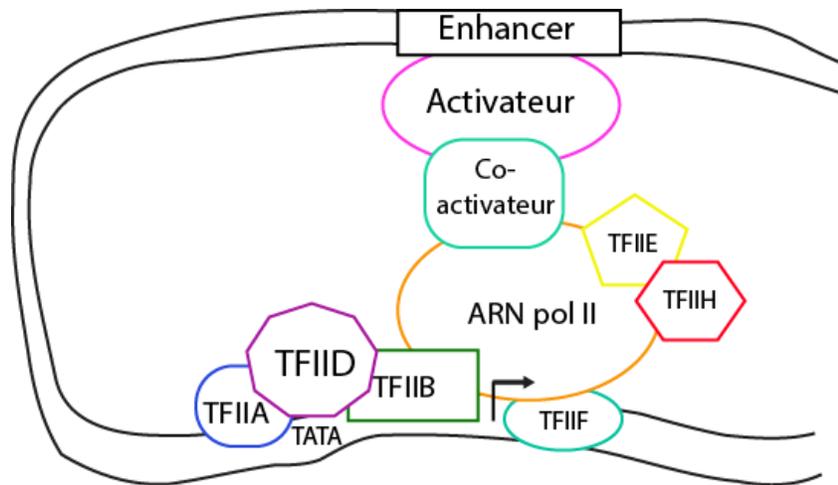


Figure 2. Schéma simplifié du phénomène d'amplification de la transcription d'un gène. Des protéines activatrices lient les enhancers et font le pont avec la machinerie basale de la transcription grâce à des co-activateurs [6, 16, 18, 19, 21].

Élongation et terminaison de la transcription

Une fois l'initiation de la transcription d'un gène entamée, il y a, de manière successive, l'avortement de transcrits naissants courts (entre 2 et 15 nucléotides) avant que l'initiation de la transcription ne soit productive [22]. Alors que l'ARN Polymérase II réussit enfin à se libérer du promoteur, débute l'étape d'élongation de la chaîne naissante de nucléotide, ayant comme matrice l'un des deux brins d'ADN. L'enzyme s'est, à ce moment-là, défaits de plusieurs des facteurs qui y étaient liés lors des étapes d'initiation [23, 24].

Le mouvement de l'ARN polymérase II, contrairement aux enzymes hélicases typiques, ayant pour fonction de séparer les deux brins d'ADN, est produit par un mouvement décrit par le cliquet Brownien. À chaque « pas », l'appariement du bon nucléotide complémentaire au brin d'ADN codant donne lieu à une réaction d'hydrolyse du nucléotide et permet à l'enzyme d'avancer. Par contre, ce genre de mécanisme laisse aussi à l'enzyme la possibilité de revenir sur ses pas [16, 25, 26]

L'ARN Polymérase II est évidemment assistée de plusieurs facteurs assurant l'élongation, tel que P-TEFb, un facteur phosphorylant la Serine en position 2 du domaine C-terminal de l'ARN Polymérase II [27, 28]. Ce domaine est aujourd'hui connu pour être phosphorylé à des positions précises selon l'étape à laquelle est rendue l'enzyme dans la transcription d'un gène. Ainsi, lors de l'initiation, c'est la Serine à la position 5 qui est phosphorylée, alors que lors de l'élongation, c'est la Serine 2 qui l'est [21, 29]. D'autres facteurs sont aussi recrutés, entre autres des protéines permettent d'ajouter la coiffe à l'extrémité 5' de l'ARNm naissant et d'autres, ont pour fonction d'épisser les introns des ARNm naissants.

Ainsi, l'ARN Polymérase II suit son cours tout en catalysant la polymérisation de l'ARNm naissant jusqu'à atteindre ce qui sera l'extrémité 3' de ce dernier. L'un des modèles proposé pour la terminaison de la transcription stipule que l'ARN Polymérase II poursuit sa trajectoire jusqu'au bout de la région codant l'ARNm jusqu'au site de polyadénylation. D'ailleurs, la terminaison de la transcription par l'ARN Polymérase II est couplée au traitement de l'extrémité 3' de l'ARNm naissant, surtout où un signal de

polyadénylation intact a été démontré comme étant un élément important dans la terminaison tant dans les cellules de levures que dans les cellules humaines [30-36]. Puis, un réarrangement allostérique aurait lieu causant le désassemblage des facteurs d'élongation et le recrutement de facteurs de terminaison de la transcription, dans un ordre encore peu connu [37, 38]. L'un des facteurs de terminaison important est TTF2, une ATPase couplant l'hydrolyse d'une molécule d'ATP avec le relâchement de l'ARN Polymérase II [39]. Les composantes sont enfin recyclées pour d'autres événements de transcription.

Ainsi, la transcription d'un gène est le résultat de plusieurs étapes successives et complexes. À chacune d'elles, l'ARN Polymérase II est accompagnée de différents facteurs assurant le bon déroulement de celles-ci.

Variabilité dans les populations isogéniques

Dans les cellules, les molécules sont souvent présentes en petit nombre faisant en sorte que les processus biochimiques ne peuvent pas être expliqués par un modèle déterministe. Ils sont plutôt caractérisés par un processus stochastique, c'est-à-dire purement aléatoire. Ainsi, au lieu de n'avoir qu'un seul scénario possible, comme dans un processus déterministe, il y en a plusieurs. Il peut s'agir d'un processus discret, ayant un nombre restreint de possibilité sans possibilité d'intermédiaire, ou non tel que le temps ou l'espace. Dans le cas de la transcription, il implique que chacun des différents facteurs mentionnés précédemment soient présents physiquement au locus du gène à transcrire. Pour mieux comprendre le caractère stochastique de processus biochimiques, il a fallu avoir la capacité d'observer ces derniers dans des cellules uniques, et de préférence en y observant des molécules uniques. Pourtant, avant les années 1990, il était impossible d'étudier la stochasticité de la transcription puisqu'avec les méthodes de biochimie traditionnelles, parce que les méthodes utilisées, telles que le buvardage de type Northern

ou même de PCR quantitative, demandent la collection des millions de cellules ce qui signifie que l'étude sera le résultat de la moyenne d'une population. C'est la venue de technique d'analyse de cellule unique au début des années 90 qui a permis de mieux investiguer la stochasticité de l'expression des gènes [40]. De telles analyses ont enfin permis de mesurer que les processus cellulaires étaient aléatoires et ont confirmé ce que plusieurs modèles théoriques ont suggéré : au sein d'une population isogénique, les cellules arborent une grande variabilité dans le nombre de transcrits et de protéines issus d'un même gène, même si ce dernier n'est pas régulé par le cycle cellulaire [41-44].

Par contre, même s'il est connu que les processus cellulaires sont stochastiques et sont sujets à des fluctuations aléatoires [43, 45], il semblait très improbable que la grande variabilité observée n'était dûe à ladite **stochasticité intrinsèque**. Celle-ci correspond à la stochasticité touchant directement la transcription. Par exemple, la proximité d'une ARN Polymérase II au site de promotion d'un gène, ou encore, la présence d'un facteur essentiel à la transcription à un endroit précis afin que cette dernière ait lieu.

La **stochasticité extrinsèque** explique aussi une partie de l'observation faite dans plusieurs laboratoires puisqu'elle est décrite comme étant issue des fluctuations d'un phénomène donné par d'autres réseaux à même la cellule [46]. Les effets du cycle cellulaire [47, 48] et des réseaux agissant en amont ont été montré comme facteurs contribuant au phénomène [48]. Plusieurs études suggèrent que ce sont les fluctuations extrinsèques qui dominent la variabilité entre les cellules dans les procaryotes [42] et les eucaryotes [44].

Corrélation entre les variations transcriptionnelles et les variations traductionnelles

À partir de l'observation selon laquelle les cellules d'une même population démontrent une variation dans le nombre de transcrits ou de protéines issus d'un certain gène, une question fondamentale fut soulevée : existe-t-il une corrélation entre la stochasticité transcriptionnelle et la stochasticité traductionnelle?

L'un des premiers groupes à tenter d'élucider la question fut celui de l'équipe de Cox. Ils ont été capables de quantifier de manière précise les fluctuations des ARN messagers (ARNm) grâce à une technique basée sur la microscopie. Elle consiste à introduire dans le génome d'une cellule, soit à l'extrémité 3' d'un gène d'intérêt, une séquence contenant 24 répétitions qui, une fois le gène transcrit, se replie en une confirmation secondaire précise reconnue par la protéine d'un phage appelée MS2. L'addition d'un plasmide contenant la séquence codante pour la protéine MS2 en fusion avec la séquence de la protéine fluorescente verte (GFP) permet l'observation en temps réel des transcrits, leur quantification ainsi que la mesure de leurs paramètres cinétiques. Dans leur étude, ils démontrent qu'il existe une corrélation entre les fluctuations transcriptionnelles et les fluctuations traductionnelles. Par contre, ils observent que ce n'est pas le cas des cellules en division [49].

Une année plus tard, le groupe de Sunney Xie a publié un article dans le journal Nature dans lequel ils expliquent que par une autre méthode basée sur la microscopie, plus précisément par utilisation un système de chambre microfluidique, ont réussi à détecter en temps réel la fluorescence émise par la protéine rapportrice β -galactosidase dans des cellules d'*Escherichia coli*. Dans leur étude, ils suggèrent que la production protéique est de fait stochastique et qu'elle serait liée à la stochasticité transcriptionnelle dans une population isogénique de cellules [50]. Par contre, ils ont utilisé une technique qui détecte bien les fluctuations protéiques, mais ne prend pas en considérations les fluctuations transcriptionnelles qui peuvent survenir dans les organismes procaryotes.

Une étude plus récente par le même groupe démontre par la technique de smFISH (voir section V. **Étude des patrons d'expression des gènes par smFISH**) et en

combinaison avec l'introduction du gène rapporteur de la protéine à fluorescence jaune (YFP) qu'il n'existe pas de corrélation au sein d'une cellule entre le niveau des ARNm et des protéines qu'ils encodent. Cette étude a été conduite dans les cellules d'*E.coli* et le groupe explique leur observation par le fait que la durée de demi-vie des transcrits dans cet organisme est beaucoup plus courte que la durée de demi-vie de ses protéines [51].

À la lumière de ces études, il est encore difficile de confirmer ou d'infirmer s'il existe réellement une corrélation entre la stochasticité transcriptionnelle et traductionnelle. Même dans les eucaryotes plus évolués, il ne semble pas nécessairement y avoir de corrélation directe entre la variabilité transcriptionnelle et traductionnelle [52]. Il sera crucial dans l'avenir d'étudier la question afin de trancher si un lien existe entre les deux ou si les deux événements sont complètement indépendants.

Patrons d'expression de gènes

Avant 2008, la cause de la variabilité d'expression d'un même gène existante entre deux cellules isogéniques restait nébuleuse. Les différents laboratoires travaillant sur le sujet se contentaient de l'expliquer par les fluctuations intrinsèque et extrinsèque. Or, cette année-là, le groupe de Robert H. Singer a démontré qu'un gène peut être transcrit par un des deux patrons d'expression qu'ils ont étudié et observé dans des cellules de *Saccharomyces cerevisiae* [53]. Cette étude a beaucoup contribué aux études dans le domaine car il s'agissait d'une des premières études démontrant qu'il existe un patron d'expression et que ce dernier est gène-spécifique dans un organisme donné et que la majeure partie de la variabilité de l'expression génique peut être expliquée par la variabilité transcriptionnelle [43-45, 48, 54, 55].

Un gène peut donc être transcrit selon l'un des deux patrons d'expression suivants : soit continuels ou par *bursting* [53]. Chacun de ces deux patrons possède des caractéristiques menant à des phénotypes d'expression de gène très distincts. Le degré de variabilité de l'expression d'un gène observé dans les cellules change donc grandement

selon s'il est exprimé continuellement ou en *burst*.

D'ailleurs, plusieurs études ont corroboré ces résultats [41, 56-61]. L'une de celles-ci était du groupe de Felix Naef qui a corroboré l'observation faite par le groupe de Robert H. Singer. En utilisant le gène rapporteur luciférase et de la modélisation mathématique, ils ont démontré qu'un gène peut posséder ses propres paramètres cinétiques [56].

i. Continuuel

Le premier patron d'expression décrit par le groupe de Robert H. Singer est le patron continuuel. Il est important de ne pas confondre ce patron d'expression, parfois appelé constitutif, avec les gènes dits constitutivement actifs. Ceux-ci sont exprimés dans différents types cellulaires et ne sont pas régulés par le cycle cellulaire, appelés *housekeeping genes* [62, 63]. Un patron d'expression définit la manière dont un gène est transcrit, entre autres la fréquence d'initiation de la transcription et la manière dont les événements d'initiation de la transcription sont reliés les uns aux autres.

Ainsi, le patron continuuel a été décrit comme étant celui qui mène à une faible variation dans le nombre de transcrits du même gène dans une population de cellules isogéniques. Dans leur étude, ils démontrent que la distribution de la fréquence du nombre de transcrits par cellule correspond à une distribution de Poisson, soit une distribution légèrement différente d'une normale. La distribution de Poisson est celle qui décrit les phénomènes aléatoires qui sont indépendants les uns des autres (Figure 3A).

Par défaut, un promoteur est inactif jusqu'à ce que le PIC y soit assemblé : c'est alors que l'ARN polymérase II peut entreprendre son activité de transcription. Ce n'est qu'à ce moment que le promoteur est considéré comme étant actif. Dans leur étude, le groupe de Robert H. Singer détermine que le patron d'expression continuuel est celui par lequel un gène est transcrit une seule fois à chaque activation du promoteur. Il n'y a donc pas de mémoire de l'initiation précédente, chacun d'elles est indépendantes de la suivante tel que le prévoit un processus stochastique. Par exemple, ils observent que la plupart des cellules

de *S. cerevisiae* transcrivent le gène *MDNI* (voir section **VI. Rationnel**) par le patron continu. C'est ainsi dire que les périodes où la transcription est inactive sont plus courtes. De plus, les fréquences de transcription observées sont très petites, où environ un ARNm est transcrit à toutes les 5-10 minutes. Puisque les complexes d'initiation sont plutôt instables, lorsqu'ils s'assemblent il semble qu'un seul ARNm soit produit.

ii. Bursting

Dans leur étude parue en 2008, Zenklusen et al. ont montré qu'un gène en particulier, *PDR5*, encodant pour une protéine transmembranaire ayant un rôle central dans le système de résistance aux drogues, montrait un différent mode d'expression. Au contraire de *MDNI*, il démontrait une distribution surtout bimodale concordant avec un processus où plusieurs transcrits sont produits dans un événement d'initiation de la transcription, allongeant donc les périodes où le promoteur est à l'état inactif, en accord avec une transcription par *burst*.

Le patron d'expression en *burst* est celui qui a été le plus observé, surtout et d'abord dans des cellules de mammifères [49, 56, 57, 59, 64, 65]. Il a aussi été observé dans des bactéries [49] et une étude parue en 2009 par Boettiger et Levine dans des embryons de drosophiles démontrait que plusieurs gènes ont comme patron d'expression le *burst*. Ils observent notamment que les noyaux exprimant ces gènes sont adjacents à des noyaux où l'on ne peut pas détecter de transcription [66]. Ainsi, il est de commun accord que ce mode d'expression pourrait être décrit par une distribution bimodale où une sous-population n'exprime presque pas un gène alors que le reste de la population l'exprime de manière accrue (Figure 3B).

Bien que ce patron d'expression a été observé à plusieurs reprises et ce, dans différents organismes, les éléments qui façonnent ses paramètres cinétiques restent inconnus.

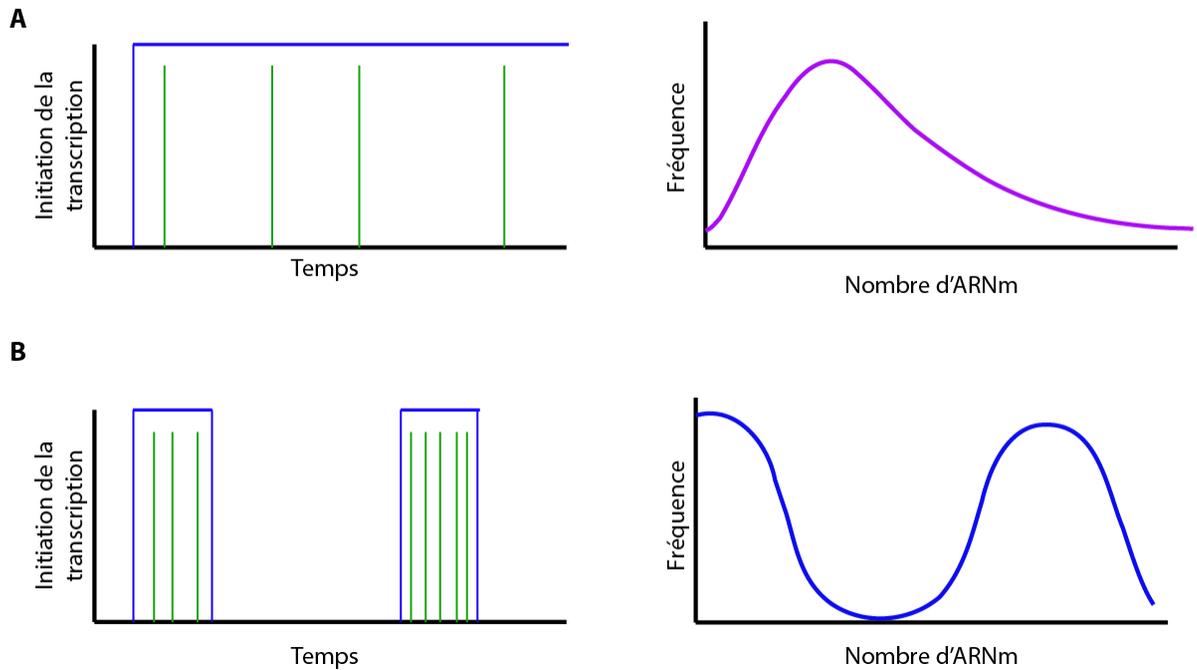


Figure 3. Patrons d'expression de gènes A. (Gauche) Représentation des évènements d'initiation de la transcription par rapport au temps pour un gène exprimé par le mode continu. En vert sont représentés les transcrits issus d'un évènement d'initiation (rectangle bleu). (Droite) Distribution de la fréquence du nombre de transcrits naissants pour un gène transcrit continuellement correspondant aussi à une distribution de Poisson. B. (Gauche) Représentation des évènements d'initiation de la transcription par rapport au temps pour un gène exprimé par le mode *bursting*. En vert sont représentés les transcrits issus d'un évènement d'initiation (rectangle bleu). (Droite) Distribution de la fréquence du nombre de transcrits naissants pour un gène transcrit en mode *burst* correspondant aussi à une distribution bi-modale.

II. Facteurs influençant l'expression des gènes

Malgré l'essor fulgurant qu'a connu le domaine depuis la publication de Robert H. Singer et son équipe, les paramètres mécanistiques et biologiques régissant les patrons d'expression de gènes présentés précédemment restent inconnus. Plusieurs facteurs ont par contre été proposés.

Contexte génomique

Il a été proposé que le contexte génomique, soit la localisation spatiale d'un gène dans le génome, pourrait dicter si ce dernier sera transcrit par l'un ou l'autre des modes d'expression ci-haut mentionnés.

Ainsi, en 2010, Skupsky et al. se sont penché sur la question en étudiant la cinétique de la transcription d'un gène rapporteur GFP sous le contrôle d'un promoteur du virus d'immunodéficience humaine (VIH). Les lentivirus, dont fait partie le VIH, intègrent de manière semi-aléatoire leur génome. Ils ont donc démontré que la localisation du gène issue de l'intégration semi-aléatoire altère la taille du *burst*, et non sa fréquence [64].

Une autre étude de 2011, utilisant le gène rapporteur de la luciférase, montre qu'un traitement à la trichostatine A, agent induisant l'hyperacétylation d'histones, a un effet significatif sur la taille du *burst* d'expression mais pas sur la cinétique globale de la transcription [56].

Deux autres études ont démontré que gènes rapporteurs insérés dans le même *locus* tendent à être exprimés simultanément, alors que l'insertion des mêmes gènes dans des *loci* différents produira une expression de ces derniers de manière aléatoire [59, 67]. Une explication est par contre donnée : l'environnement génomique dans lequel est placé un gène déterminera le paramètre temporel des *bursts* ce qui contribue à une recrudescence dans la synchronisation avec un autre gène placé dans le même *locus*. Il a aussi été proposé

que cela pourrait être causé par un recrutement synchronisé des dites factoreries de transcription, compartiments spécialisés et riches en ARN Polymérase II [68].

Ces études démontrent que le contexte génomique contribue à une partie des paramètres cinétiques de la transcription, mais de manière limitée. Ainsi, il ne s'agit pas du phénomène principal dictant qu'un gène aura un patron d'expression continuels ou en *burst*.

Pause de l'ARN Polymérase II

Un deuxième phénomène étudié comme ayant une contribution importante aux paramètres cinétiques de la transcription d'un gène est la pause que fait l'ARN Polymérase II après l'initiation, mais juste avant le début de l'élongation (Figure 4).

Longtemps, la pause de l'ARN Polymérase II sur un gène a été vue comme étant un facteur limitant à la transcription de ce dernier [27]. Pourtant, les récentes études de génome entier, tel que le projet ENCODE [69], ont plutôt démontré que la pause de l'ARN Polymérase II est beaucoup plus fréquente que prédite. Ce phénomène a été surtout observé dans des embryons en développement de *Drosophila melanogaster* et dans des cellules de mammifères précédant l'activation de la transcription d'un gène [27, 70-72].

Dans les embryons de *D. melanogaster*, l'étude de Boettiger et Levine ayant été menée dans quelques gènes dans le développement en utilisant la méthode traditionnelle d'hybridation fluorescente *in situ* (FISH). Cette étude démontre que les gènes où l'on observait une pause de l'ARN Polymérase II au promoteur avait un patron d'expression continuels, alors que les gènes où l'on n'observait pas la pause de l'enzyme avaient plutôt un patron en *burst* [66]. Ils prédisent donc que la plupart des gènes impliqués dans l'axe dorso-ventral, qu'ils ont en partie étudiés, contiennent l'ARN Polymérase II en pause et que ces derniers devraient avoir un patron d'induction de la transcription continuels.

Bien que cette étude représente une nouvelle avenue vers la compréhension de la cause menant aux différents modes de transcription, l'hypothèse ci-haut suggérée n'a pas été prouvée avec une méthode pouvant détecter de manière précise et quantitative les transcrits naissants dans les cellules. En effet, dans l'étude de Boettiger et Levine, ils

mentionnent qu'ils ont détectés une partie des transcrits naissants dans les embryons, mais pas tous.

Enfin, il a été démontré que les données issues d'expériences de ChIP-chip montrant la présence de l'ARN Polymérase II au promoteur d'un gène ne prouvent pas qu'elle y fait une pause. En fait, pour en être complètement assuré il faut procéder à un essai de promoter melting au KMnO_4 en plus du ChIP-chip. Cet essai consiste à ajouter ce composé chimique à des cellules intactes qui auront des résidus entre 20 et 50 paires de bases correspondant grandement aux régions où l'ARN Polymérase II fait une pause [73]. Un autre essai devrait être fait pour valider la présence de l'ARN Polymérase II au site de promotion d'un gène : Immunoprécipiter la chromatine (ChIP) en suivant la présence du facteur d'élongation négatif (NELF), puisque ce dernier est connu comme étant un répresseur de l'élongation de la transcription et permet une pause plus efficace de l'ARN Polymérase II [13].

À la lumière de ces études, la présence de l'ARN Polymérase II en pause au promoteur d'un gène n'est pas nécessairement un facteur déterminant pour le mode d'expression que ce gène en question arborera. Une corrélation existe entre les deux, mais d'après les études menées à ce jour, il ne s'agit pas d'une relation de cause à effet.

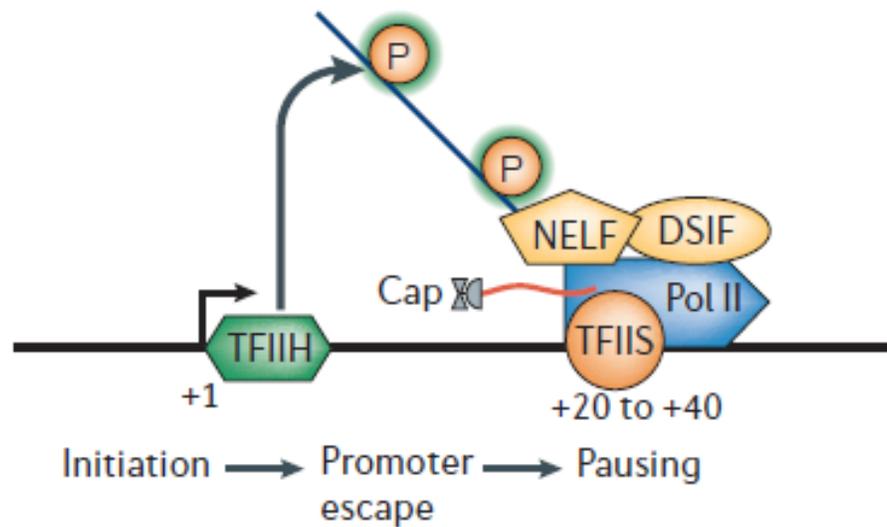


Figure 4. Représentation de la pause de l'ARN Polymérase II après l'initiation de la transcription. De manière générale, les facteurs NELF (Negative Elongation Factor) et DSIF (DRB Sensitive Inducing Factor) sont recrutés et s'associent à l'enzyme, l'empêchant de poursuivre son activité. Une série de phosphorylations et de déphosphorylations par d'autres facteurs permettent la reprise de l'activité de l'ARN Polymérase II. (Tiré de [27])

Séquence et architecture du promoteur

La séquence et l'architecture du promoteur est un autre facteur pouvant être impliqué dans la régulation des patrons d'expression de gènes ou dans la stochasticité de leur expression.

En 2010, un article de l'équipe de Shav Tal a été publié dans lequel ils réussissaient à quantifier de manière précise des transcrits naissants dans des cellules humaines de type HEK293 en temps réel [74]. Ils ont d'abord recombinaison la copie unique de la région codante du gène de la cycline D1 soit sous le contrôle de son propre promoteur ou par le promoteur du cytomégalo virus (CMV) avec des régions répétées dans la partie du 3'UTR. Une fois transcrites, les répétitions se trouvant dans la région 3'UTR du gène sont reconnues par une protéine de bactériophage MS2 en fusion avec la protéine à fluorescence verte (GFP). Ainsi, les ARNm sont faciles détectables, mais surtout, il est simple de les suivre en temps réel. Cette technique permet de quantifier de manière spécifique les transcrits du gène de la Cycline D1. Dans leur étude, ils observent un changement de patron d'expression lorsque le promoteur original du gène de la Cycline D1 est changé pour celui du CMV. Alors que le gène sous le contrôle de son promoteur natif a une expression en *burst*, celui du CMV lui confèrerait une expression continue. Il a été clair, à la lumière de cette étude que le promoteur d'un gène est l'élément déterminant à la manière dont un gène est exprimé.

Zenklusen et al. ont montré en 2008 que le gène PDR5 dans la levure *S. cerevisiae*, qui avait un patron d'expression en *burst*, possède en sa région promotrice une boîte TATA, élément pouvant moduler le patron d'expression du gène [53].

Dans la même voie, le travail de l'équipe de Naef avait aussi montré la possibilité d'un rôle important de ces éléments dans le mode d'expression d'un gène. En utilisant le gène rapporteur de la luciférase, ils ont étudié l'expression d'un gène avec différentes régions promotrices, soit avec une boîte TATA et une ou plusieurs boîtes CAAT. Ils concluent que le gène contenant la boîte TATA présente une expression en *burst* ainsi que lorsque le nombre de boîtes CAAT est élevé. Par contre, cette étude relève les aspects de *burst* traductionnel plutôt que transcriptionnel, ce qui nous intéresse.

Des études à haut débit du groupe Weissman, où l'expression des protéines de la levure à bourgeonnement a été étudiée grâce à leur collection GFP et à l'analyse par cytométrie en flux, a suggéré que les gènes contenant une boîte TATA en leur promoteur étaient plus "bruyants" en terme d'expression [75]. Cela indique donc que la boîte TATA détient un rôle dans le mode d'expression d'un gène [44, 75]. D'autres études, où la boîte TATA a été mutée, ont corroboré cette prémisse [54, 76, 77].

À ce jour, les études démontrent qu'il existe un lien entre l'architecture générale du promoteur, telle que la présence d'une boîte TATA, et la régulation de l'expression génique. Des études plus approfondies sont nécessaires pour tenter de comprendre l'impact réel des séquences et de l'architecture d'un promoteur sur le patron d'expression du gène qu'il contrôle.

Gene Looping

Une fois la transcription terminée, l'ARN Polymérase II rencontre un signal de terminaison où elle fait une pause avant d'être relâchée de l'ADN. Par contre, il a été démontré que pour certains gènes, l'ARN Polymérase II peut reprendre un nouveau cycle de transcription grâce aux *gene loops*, rapprochement physique entre le promoteur et les régions terminatrices d'un gène [78] (Figure 5).

Les *gene loops* ont été découvert dans plusieurs organismes à différents stades évolutifs : en commençant par le phage λ [79], jusqu'aux mammifères [80, 81] en passant par la levure [78] et même chez le virus de l'immunodéficience humaine [82].

Bien qu'aucune étude ne fait le lien direct entre les *gene loops* et le mode de transcription en *burst*, il n'en reste pas moins une hypothèse que certains croient plausible puisque ces *gene loops* favoriseraient la transcription du même gène plusieurs fois en un même cycle d'initiation [83].

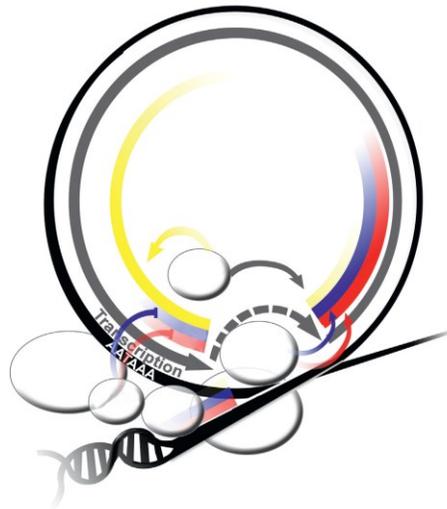


Figure 5. Représentation schématique du phénomène de ré-initiation de la transcription par *gene looping*. Une boucle est formée entre la région promotrice et l'extrémité 3' d'un gène permettant à l'ARN Polymérase II de réinitier la transcription [83].

III. Évolution de la complexité des génomes

La caractéristique qui sépare le plus drastiquement les génomes des procaryotes des génomes d'eucaryotes est la taille du génome (des régions non codantes) et la présence d'introns qui sont épissés dans ces derniers. Pourtant, la quantité et la densité des introns diffèrent énormément au sein des organismes eucaryotes. Par exemple, certains organismes unicellulaires ont un ou deux introns dans tout leur génome, alors que la plupart des vertébrés ont en moyenne entre cinq et huit introns par gène [84-86]. Même dans la taille ceux-ci diffèrent : alors que la plupart des introns ont entre 20 et 200 nucléotides chez la levure, les introns retrouvés dans le génome humain ont une moyenne de 2 kb [87].

La levure *Saccharomyces cerevisiae*, l'un des modèles eucaryotes le plus utilisé en biologie moléculaire, aussi connue sous le nom de levure à bourgeonnement, possède un génome d'environ 12 millions de paires de bases réparties sur 16 chromosomes et ayant une densité d'environ 483 gènes répartis par millions de bases. Contenant peu d'introns (~250), cet organisme possède un génome compact [88, 89]. L'avantage d'étudier cet organisme réside dans le fait que ce dernier a une structure génomique ainsi qu'une architecture des régions promotrices des gènes plutôt simples.

La levure à fission, *Schizosaccharomyces pombe*, possède quant à elle un génome comparable à celui de *S. cerevisiae* avec 14.1 Mb réparties sur 3 chromosomes. Quoique son génome soit aussi similaire en densité à *S. cerevisiae* avec environ 345 gènes par millions de base, *S. pombe* contient un grand nombre d'introns (~5000) et est évolutivement plus complexe que son homologue [88, 89]. De plus, cet organisme se divise dans son axe médial contrairement au bourgeonnement de *S. cerevisiae* [90, 91]. Il s'agit donc d'un modèle intéressant à étudier, mais surtout, facile à étudier lors des différentes phases de son cycle-cellulaire. D'ailleurs, tout comme *S. cerevisiae*, cet organisme a aussi été longtemps étudié, mais surtout pour son cycle cellulaire.

Le vers *Caenorhabditis elegans* possède un génome comportant 100 millions de

paires de bases sur 6 chromosomes [92]. Ce dernier organisme possède une densité d'environ 200 gènes par millions de bases, ce qui est plutôt similaire à *S. pombe* tout en étant un organisme eucaryote assez complexe. Il s'agit déjà d'un organisme modèle dans plusieurs laboratoires.

La mouche à fruits, ou *Drosophila melanogaster*, est un organisme complexe étudié surtout pour son cycle de développement. Possédant quatre paires de chromosomes contenant en tout 169 Mb, le génome de cet organisme a une densité d'environ 86 gènes par Mb [92, 93]. Ainsi, il a une organisation plus complexe et moins compacte.

Ultime modèle étudié, l'humain est un organisme très complexe avec 3137 Mb réparties sur 23 paires de chromosomes et ayant environ 7 gènes par Mb [92]. Cela signifie que le nombre de régions intergéniques est très grand et que son génome est amplement complexe. Pour étudier ce dernier organisme, il est possible d'utiliser des lignées cellulaires cancéreuses, telles que les cellules HEK293T et U2OS, puisque ces dernières sont faciles à cultiver en laboratoire.

Ainsi, plusieurs modèles sont disponibles pour des fins d'études à différents stades évolutifs et de complexité.

IV. Patrons d'expression des gènes dans l'évolution

Patron continu chez les organismes eucaryotes inférieurs

Dans les dernières années, les études sur les patrons d'expression de gènes se sont multipliées permettant de mieux comprendre ces derniers et surtout d'étudier ce qui les influence au sein d'une cellule [42, 43, 56, 57, 59, 61, 66, 94, 95]. Certains gènes ont été caractérisés comme étant transcrits continuellement, alors que d'autres par *burst*. Il est possible d'observer, à la lumière de ces études, que le patron d'expression continu a été observé pour plusieurs gènes dans la levure *S. cerevisiae*, mais rarement dans des organismes eucaryotes supérieurs [53, 66]. Il a été proposé que cette différence pourrait être due au fait que les eucaryotes inférieurs ont un cycle de vie plus court que les eucaryotes supérieurs, ainsi leurs gènes requerraient d'être transcrits continuellement. De plus, la simplicité de leurs promoteurs permettraient à la machinerie transcriptionnelle d'être assemblée plus facilement que sur les promoteurs d'eucaryotes supérieurs [53, 96].

Patron *bursting* chez les organismes eucaryotes supérieurs

Bien que le gène de *PDR5* dans la levure *S. cerevisiae* possède un patron d'expression en *burst*, peu de gènes contenant une boîte TATA ont été investigués. À ce jour, le patron en *burst* a surtout été observé dans les organismes eucaryotes supérieurs tels que les *D. discoideum*, *D. melanogaster* et *H. sapiens* [43, 56, 57, 59, 65, 66, 95]. Même si quelques gènes dans la mouche à fruits possèdent un patron d'expression en continu [66], la majorité des gènes étudiés dans les eucaryotes supérieurs ont un mode d'expression en *burst*.

Il a été suggéré que le patron d'expression en *burst* jouerait un rôle important dans le développement et dans la différenciation cellulaire d'organismes complexes [97-100]. Par exemple, dans la mouche à fruits, deux protéines, Notch et Delta, ont des fluctuations stochastiques et cela semble être un facteur important dans l'émergence des cellules

précurseurs neuronales à partir d'une population isogéniques de cellules [101]. Chez l'humain, ce patron a aussi été montré comme étant important dans la différenciation cellulaire [102]. Dans ces exemples, l'importance de la stochasticité réside dans l'établissement de la sélection et de la propagation de l'expression de gènes type-cellulaire spécifiques et donc, ultimement, la différenciation des cellules.

V. Étude des patrons d'expression de gènes par smFISH

Dans les années 1980, les techniques de visualisation des composantes d'une cellule étaient limitées à l'utilisation de radio-isotopes ou encore à des particules d'or. Même avec la venue de techniques telles que les micropuces ou les buvardages de type Northern, aucun ne permettait l'observation de l'ADN ou de l'ARN de manière spécifique dans les cellules.

Ainsi, la première utilisation de la fluorescence pour l'observation d'ADN avec des sondes d'ARN spécifiques ayant un fluorophore à leur extrémité 3' a été élaborée par l'équipe Bauman puis par Singer et Ward [103, 104]. Les différentes variantes de cette technique se sont multipliées afin d'augmenter la spécificité des molécules observées et le ratio signal/bruit de fond obtenu en microscopie. De plus, les caméras captant les différentes images de microscopie en fluorescence se sont aussi ajustées en améliorant constamment la résolution des images. Enfin, les méthodes indirectes de détection, par anticorps couplés à des fluorophores détectant les oligonucléotides par exemple, ont été utilisées afin d'accroître le signal.

En 1998, le groupe de Robert H. Singer a utilisé des oligonucléotides spécifiques d'environ 50 bases portant cinq fluorophores par oligonucléotide [105]. Différentes modifications de ce protocole ont été faites, par exemple, l'équipe de Tyagi utilise des oligonucléotides plus courts, environ 20 bases, mais marqués une seule fois plutôt que cinq [95, 106] (Figure 6A). Similairement aux 50-mers, l'utilisation de 48 de ces oligonucléotides libellés une seule fois crée un signal spécifique à la détection d'ARNm uniques ainsi que des ARNm présents au site de transcription (Figure 6B).

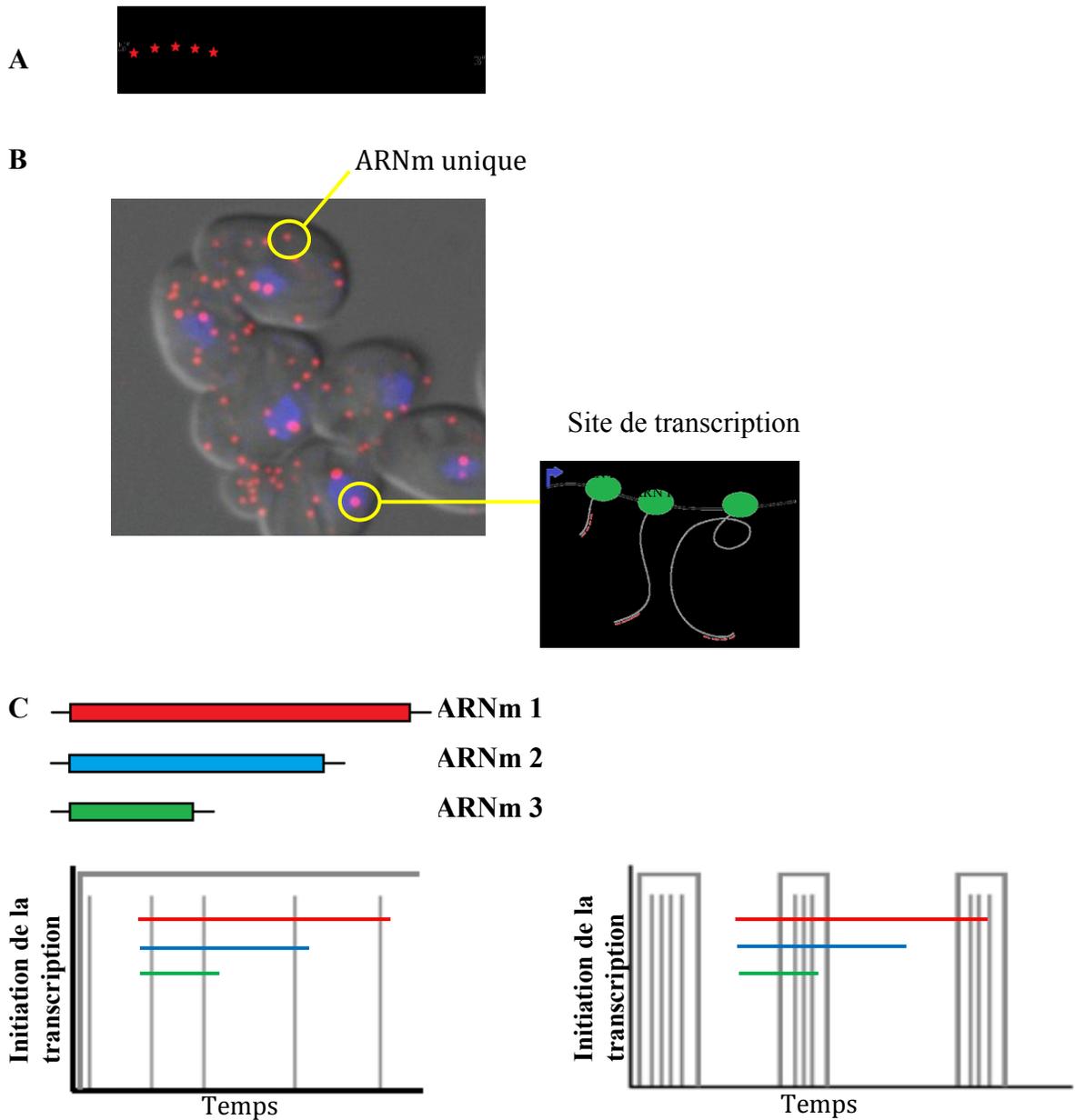


Figure 6. **A.** Représentation schématique des sondes fluorescentes utilisées dans la technique de smFISH hybridées à un ARNm spécifique. **B.** Image superposée obtenue par smFISH de cellules de *S. cerevisiae* hybridées par des sondes ciblant l'ARNm de *mdn1*⁺ en rouge, le noyau visible grâce au DAPI et la forme des cellules par lumière transmise. Les cercles jaunes entourent un ARNm unique de *mdn1*⁺ ainsi qu'un site de transcription où l'accumulation de la fluorescence des sondes hybridées aux transcrits naissants. **C.** Représentation de trois ARNm de longueurs différentes ainsi que la probabilité d'en observer des transcrits naissants et l'intensité relative supposée être observée selon le mode d'expression employé.

Depuis, cette variante est devenue très populaire et permet la détection et la quantification de molécules uniques d'ARN dans n'importe quelle lignée cellulaire fixée sans même avoir à introduire des modifications génétiques dans cette dernière [95, 105-108].

Enfin, cette technique permet d'avoir de l'information sur l'initiation de la transcription de gènes ayant différentes longueur. En fait, un gène très long exprimé continuellement sera visible dans presque toutes les cellules au site de transcription. Si ce dernier est exprimé en *burst*, les sites de transcription seront visibles dans presque toutes les cellules et contiendront un très grand nombre de transcrits naissants de ce gène. Cela s'explique par le simple fait que ce dernier prendra plus de temps à être transcrit. Plus le gène est court, moins souvent il sera possible d'observer un site de transcription pour ce dernier peu importe son patron d'expression (Figure 6C). Ainsi, le nombre de transcrits naissants par site de transcription quantifiables grâce au smFISH déterminera son mode d'expression.

VI. Rationnel

Il a été démontré qu'un gène peut être exprimé soit par le patron d'expression continuuel ou par *bursting* [53]. Les différentes études menées à ce jour ont montré que la plupart des gènes étudiés dans des eucaryotes inférieurs, telle que la levure *S. cerevisiae*, sont exprimés selon le patron d'expression continuuel, alors que les gènes des eucaryotes supérieurs ont plutôt un patron d'expression en *burst* [53, 56, 57, 59]. Il est possible que la complexité des eucaryotes supérieurs favorise le mode d'expression en *burst*. Par contre, aucune étude n'a pu démontrer cela. Les différents groupes travaillant sur ce sujet observent quelques gènes dans un organisme de leur choix, mais aucune étude systématique n'a été faite.

Le but de ce projet est donc d'étudier par smFISH le patron d'expression de quatre gènes (*mdn1*⁺, *PRP8/spp42*⁺, *pol1*⁺ et *cdc13*⁺) évolutivement conservés et présents en une seule copie par génome, afin de faciliter la quantification par smFISH, dans plusieurs organismes à des stades évolutifs différents.

Nous étudierons les patrons de transcription dans différents organismes, chacun reflétant différente complexité, en particulier par rapport à leur complexité génomique. Nous étudierons deux organismes eucaryotes simples, *S. cerevisiae* et *S. pombe*, ainsi que trois organismes eucaryotes supérieurs : *C. elegans*, *D. melanogaster* et des cellules humaines, ultime modèle d'organisme eucaryote complexe.

B. Materials and Methods

1. Single-molecule resolution Fluorescent *in situ* Hybridization (smFISH)

Single-molecule resolution FISH was performed following most of the protocol previously published by Zenklusen D. and Singer R.H., and later by Zenklusen D. and Rahman S. [107, 109].

1.1 Probe design and labeling

The probe sets (BioSearch Technologies) of each studied gene, for all organisms are chosen using the Stellaris Probes Designer (BioSearch Technologies) with a GC content higher than 40%.

Labelling of probes is achieved by first purifying a total of 10 μ g of each gene's probe mix with the QIAquick Nucleotide Removal Column (QIAGEN), following the manufacturer's instructions. Elution must be done using water, followed by lyophilisation of purified probe mixes in a SpeedVac. We added 30 μ L of labelling solution (0.1M sodium bicarbonate, pH 9.0) to each 50 μ g tubes of either DyLight 550 or 594 (Thermo Fisher Scientific) and then added 15 μ L of the labelling-solution into each lyophilised probe sets. Each DyLight tube was used to label two probe sets.

The labelling efficiency was calculated for each set of probes using equations 1-3 following the measurement of absorption by a NanoDrop (Thermo Fisher Scientific).

$$\text{(Equation 1)} \quad [\text{DNA}] = \frac{A_{\text{DNA}} - \epsilon_{\text{dye}(260)} \times (A_{\text{dye}} / \epsilon_{\text{dye}(\text{max})})}{\epsilon_{\text{DNA}} \times 0.1 \text{ cm}}$$

$$\text{(Equation 2)} \quad [\text{Dye}] = \frac{A_{\text{dye}(\text{max})}}{\epsilon_{\text{dye}} \times 0.1 \text{ cm}}$$

$$\text{(Equation 3)} \quad \text{Labeling efficiency} = \frac{[\text{Dye}]}{[\text{DNA}]}$$

Equation 1 allows to calculate the exact DNA concentration contained in each tube, equation 2 allows to calculate the exact concentration of the dye and equation 3 calculates the labelling efficiency. In all equations, A_{DNA} corresponds to the absorption of the labelled oligonucleotides at 260nm, $\epsilon_{\text{dye}}(260)$ corresponds to the extinction coefficient of the dye at 260nm ($\epsilon_{\text{dye}}(260)$ of DyLight 550 is 12150 and of DyLight 594 is 46800), A_{dye} corresponds to the maximum measured absorption of the dye at its respective emission wavelength, $\epsilon_{\text{dye}}(\text{max})$ is the maximum extinction coefficient of each dye (150000 for DyLight 550 and 80000 for DyLight 594).

1.2a In *Schizosaccharomyces pombe*

The wild-type *S.pombe* strain 972 h- is grown in rich medium containing 6% (w/v) of yeast extract (Bio Basic inc), 2% glucose (v/v), 15% (w/v) Adenine (Sigma Aldrich). The cells are grown at 30°C until reaching an optical density of 0.6 corresponding to the mid log-phase in *S.pombe*. Once harvested, the cells are fixed in 4% (v/v) EM grade paraformaldehyde (Electron Microscopy Sciences) for 40 minutes. The cells are washed three times in Buffer B (1.2M Sorbitol, 100mM KHPO₄ pH 7.5) before being digested in a

final concentration of 320 μ g/ μ L lysing enzymes from *Trichoderma harzanium* (Sigma Aldrich) containing 1.2 M Sorbitol, 100mM KHPO₄ pH 7.5, 20mM ribonucleoside-vanadyl complex (New England Biosearch) and 20mM b-mercaptoethanol. Cells are then washed three times with Buffer B and resuspended in 0.001% Triton X-100 (EMD chemicals)/1X PBS and placed on a shaker for 20 minutes to permeabilize cell walls. Cells are washed three more times with Buffer B and once with 10% Formamide/2X SSC solution prior to overnight hybridization at 37°C with 25ng of singly-labeled with a Dyelight594 fluorophore 20-mer oligonucleotides (for *mdn1*⁺, *prp8*, *pol1*⁺ or *cdc13*⁺) and 20ug of carrier *E.coli* tRNA/ssDNA (1:1). The final labeling solution contains 10% formamide (v/v), 0.5% NaHPO₄ pH 7.5 (v/v), 2X Saline-sodium citrate solution (300mM NaCl, 30mM Na₃Citrate·2H₂O), 2ug/uL BSA (NEB) and 10mM VRC. The cells are then washed in 10% Formamide/2XSSC, then in 0.1% Triton X-100/2XSSC and finally in 1X PBS. The cells are then placed on poly L-Lysine coated coverslips and mounted on slides with the liquid ProLong® antifade DAPI-containing reagent (Invitrogen).

1.2b In human cell-lines

Two cell-lines were used to perform smFISH experiments: U2OS and HEK293T. The IMR90 cells were a generously given to us, fixed, by Gerardo Ferbeyre's laboratory. The cell-lines were respectively grown in McCoy 10% Fetal Bovine Serum (FBS) and DMEM 10% FBS. The cells were put on poly L-Lysine coated coverslips when reached an 80% confluence and incubated 24 hours at 37°C. The cells were then fixed for 30 minutes with a 4% PFA/ 1X PBS solution and washed three times with Buffer B. The cells are then placed in 70% ethanol for cell membrane permeabilization and incubated at -20°C overnight. Cells are then rehydrated twice with 2mL of 2X SSC at room temperature for 5 minutes. A last wash with 10% Formamide/2X SSC is performed prior to probe overnight hybridization. This step was performed at 37°C with 25ng of singly-labeled 20-mer oligonucleotides of *MDNI* (5' exons probe set labelled with DyeLight 550 and 3' exons probe set labelled with DyeLight 594) and 20ug of carrier *E.coli* tRNA/ssDNA (1:1). The final labeling solution contains 10% formamide (v/v), 0.5% NaHPO₄ pH 7.5 (v/v), 2XSSC,

2 μ g/ μ L BSA (NEB) and 10mM VRC. The cells were then washed in 10% Formamide/2XSSC, then in 0.1% Triton X-100/2XSSC and finally in 1X PBS. The coverslips were finally mounted on slides with the liquid ProLong® antifade DAPI-containing reagent (Invitrogen).

2. smFISH images acquisition and analysis

The image acquisition of both organisms was performed with the same upright Carl Zeiss Axio Imager Z2 epifluorescence microscope. A 100X oil objective (Numerical Aperture =1.4) was used for *S. pombe* acquiring, and a 63X oil objective (NA=1.4) was used to acquire the human cell-lines images.

The filter cubes used to perform image acquisition were common to both organisms: a DAPI specific filter cube for nuclear signals, green channel filter cube for *S. pombe* autofluorescence for cell segmentation, a custom narrow Dyelight 550 filter cube (Emission: 546nm, Excitation: 567 nm \pm 7.5) and a custom Dyelight 594 filter cube (Emission: 581nm, Excitation: 617 nm \pm 20). The images were taken with the Carl Zeiss AxioCam MRm at a 1388 x 1040 pixels resolution and a pixel size of 6.45 μ m (h) x 6.45 μ m (v). Z-stacks were taken at every 0.2 μ m and vary according to the field taken.

2.1 *Schizosaccharomyces pombe* analysis

The images of each field were maximum projected (DAPI, green channel and DyeLight594). The cells on each field were segmented using CellProfiler [110, 111], using the green channel image for the cellular edges segmentation and the DAPI image for the nuclear segmentation. The thresholding method used to segment cells is the RidlerCalvard PerObject and the one used for the nuclei is the Kapur PerObject.

Localize, developed by Daniel Larson (NIH), is an IDL based script that was used to detect each spot in the maximum projected image of smFISH for each gene. Each spot's intensity was determined by a fit of a Gaussian mask. The detected spot's xy coordinates

and intensity were exported in a text file used for the final quantification in ARNQuant on MATLAB. ARNQuant, developed by Emmanuel Noutahi, allows to merge the segmented cells and nuclei with the detected spots from Localize. The intensity of a single mRNA is calculated from the following equation:

$$w_i = \text{round} \left(\left(\frac{\textit{Spot intensity}}{\sum \textit{cyto. spot intensity}} \right) * \textit{Number of cyto. spots} \right)$$

As the segmentation of the nucleus may be slightly off, spots that are on the edge of the nucleus are counted as a transcription site if there are less than two spots in the nucleus, as we assume there could be at most two transcription sites per nucleus. If there are more than two spots in the nucleus, the two brightest spots are considered as the two transcription sites and the remaining are considered cytoplasmic.

Cell-cycle stages quantification. In order to quantify the transcripts per cell at different cell-cycle stages, another MATLAB based script was developed by Emmanuel Noutahi. MergeQuant allows to merge dividing cells and categorize the data in three different phases: G2, M and G1/S. As the G1/S phase occurs prior to cytokinesis, the septation, visible both under transmitted light and in the green channel because of autofluorescence, is the indicator of the beginning of G1/S phase in dividing cells. The categorization is done manually, while MergeQuant combines the number of nascent and cytoplasmic transcripts of merged cells and calculates the distance between the center of the nuclei and total length of the dividing cells.

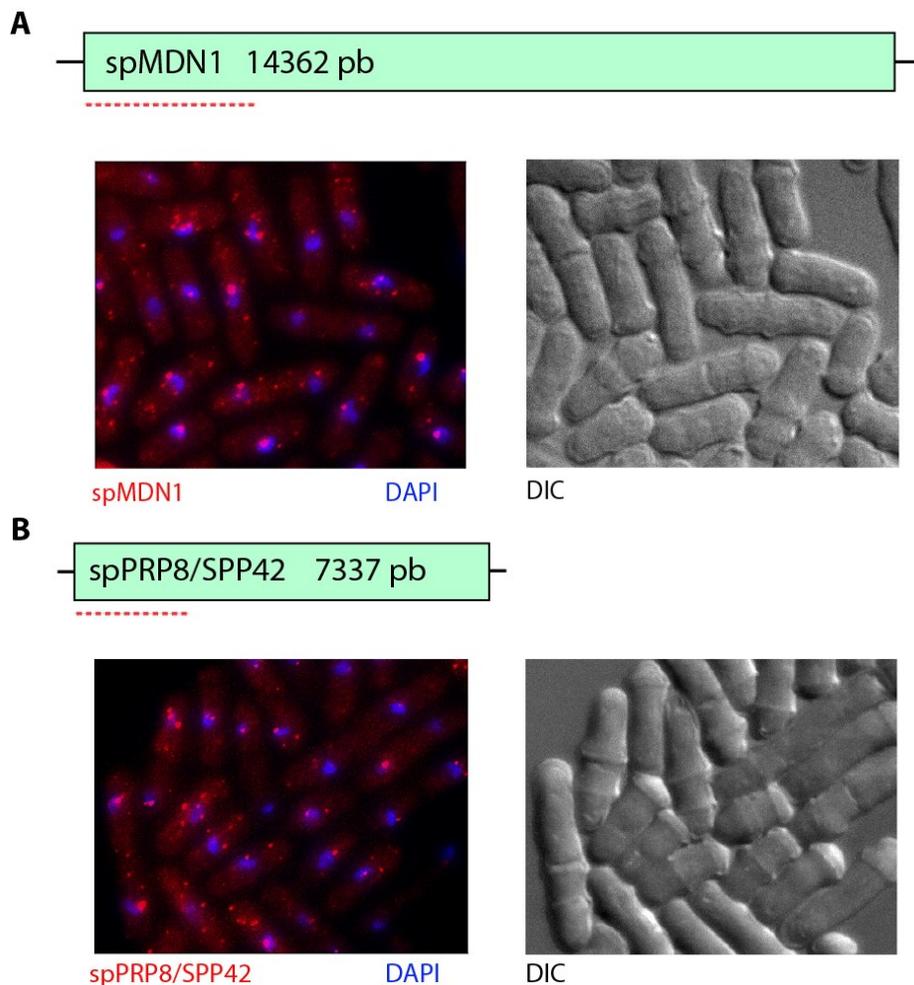
C. Results

Establishing smFISH in *Schizosaccharomyces pombe*

It has been previously shown that genes studied in lower eukaryotes often display a continuous pattern of expression that can be explained by a simple Poisson process. However, most genes studied in higher eukaryotes were shown to be expressed in a *bursting* manner, resulting in a bimodal distribution of mRNA expression. Thus far, no systematic study has been assessed to comprehend if there is a direct link between the pattern of expression of a gene and the organism in which it is expressed. This project aims to be a starting point in this systematic study. By studying evolutionarily conserved genes in *S. pombe* and in human cell-lines, and comparing with results obtained previously in *S. cerevisiae* by Zenklusen et al., we first needed to establish the smFISH technique in *S. pombe*, which has not been previously done.

With the aim of assessing the expression pattern of each gene in *S. pombe*, we first had to establish smFISH in this organism. Many changes were required modifying the pre-established smFISH protocol in *S. cerevisiae*, starting with the wall-digesting enzymes. In *S. cerevisiae* smFISH uses lyticase, however, *S. pombe* contains a different wall structure and lyticase did not efficiently digest pombe cell walls. We therefore tested different concentration of zymolyase and lysing enzymes from *Trichoderma harzanium* to have cells digested enough for the smFISH to work without over digesting the cells. Although we achieved good digestion conditions, the signals detected were not strong enough for quantification and we had to increase the number of probes we used per RNA. In *S. cerevisiae*, typically 48 probes were used to detect single mRNAs while we used 60 probes for *mdn1*⁺, *PRP8/spp42*⁺ and *pol1*⁺. We were only able to use 42 probes for *cdc13*⁺ as the RNA is small and does not allow the hybridization of more probes. Furthermore, we also

used a higher concentration of probes per hybridization, (20 μg of each probe set per reaction). At last, we changed the hybridization conditions and performed the hybridizations in liquid compared to the hybridization on coverslips, resulting in a higher density of cells per field, allowing more efficient quantification of large number of cells. All these changes allowed to detect single mRNAs for all the genes we study (Figure 7). As shown in Figure 7 and similar to what is observed in *S. cerevisiae*, we detect bright nuclear signals corresponding to sites of transcription while the single mRNAs are observed in the cytoplasm show a more regular signal intensity.



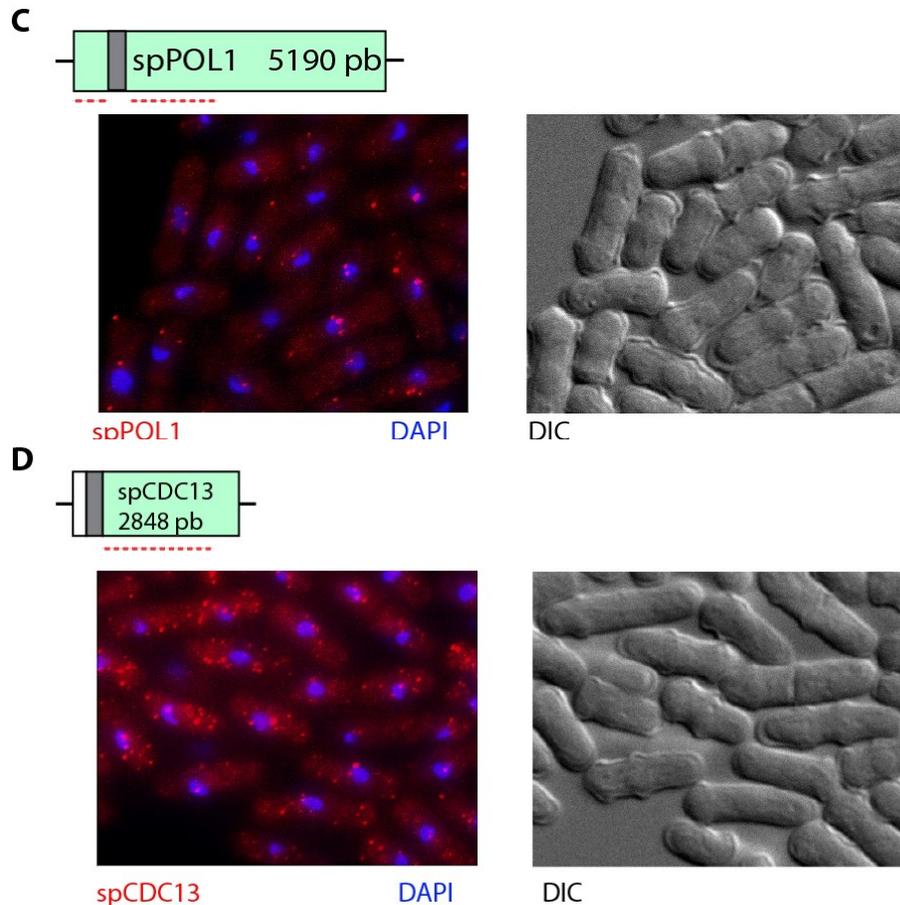


Figure 7. Graphic representation of each gene's length and structure, introns are represented in dark grey and coding regions in light green [112-114]. Left: smFISH merged image of nuclear signal in blue (DAPI) and each gene in red (Cy3.5) for *mdn1*⁺ (A), *PRP8/spp42*⁺ (B), *pol1*⁺ (C) and *cdc13*⁺(D). Right: cells shown on left panel under transmitted light.

Developing 2D image analysis tools in *Schizosaccharomyces pombe*

Previously our lab has used a combination of analysis programs to detect and quantify mRNA signal as well as to segment cellular and nuclear boundaries. However, cellular segmentation was done manually. Furthermore, integration of all data was done by copying data from different programs into Microsoft Excel. To allow more efficient quantification of your data, we developed new analysis tool allowing automated cell segmentation and data integration. First, we searched for a program allowing automated cell segmentation and chose to use Cellprofiler, a freeware program developed at the Broad Institute that allows to segment cells and nuclei. Nevertheless, the segmented cells and

nuclei output from CellProfiler are not compatible with the IDL-based scripts developed to integrate cell segmentation and mRNA quantification in the past. Emmanuel Noutahi, a summer intern in the laboratory, therefore developed MATLAB-based scripts to merge and compile the data of segmented cells and nuclei with the data of detected spots from Localize (Figure 8).

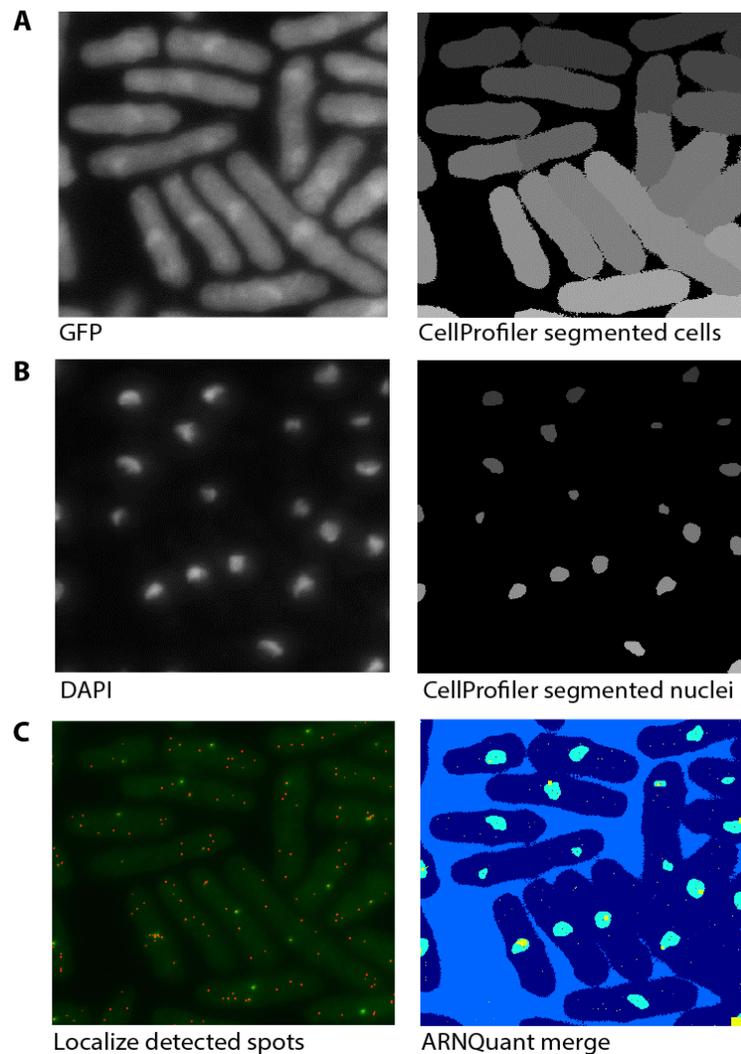


Figure 8. Cells segmented by Cell Profiler (A-right) using autofluorescence in GFP channel (A-left) in an epifluorescence microscope [110]. Nuclei segmented by Cell Profiler (B-right) using DAPI signal (B-left) [110]. Spots are detected by Localize using 2D Gaussian fit (C-left). ARNQuant, a MATLAB based script (developed by Emmanuel Noutahi) merging of segmented cells, segmented nuclei (C-right) [53, 115].

mdn1*⁺ is expressed in a continuous-like fashion in *S.pombe

The first gene studied is *mdn1*⁺, encoding for the protein Midasin. The latter is an essential AAA-ATPase for maturation and nuclear export of the pre-60S ribosomal subunit [116]. The expression of *mdn1*⁺ was studied in *S. cerevisiae* and displayed a continuous pattern of expression [53]. We chose this gene because it is a housekeeping gene, constitutively active in all organisms and cell-types and evolutionarily conserved. Its expression does not depend on phases of the cell-cycle.

In their article in 2008, Zenklusen et al. have suggested that the most precise way to determine the expression pattern of a gene is the distribution of nascent transcripts. In their article, as well as in the presented project, the nascent transcripts correspond to the detected spots within each nucleus (Figure 7) and correspond to the number of polymerases present on the studied gene at the moment of fixation (see details in **Section B. Material and Methods**).

We first determined the expression pattern of *mdn1*⁺ in *S. pombe* using smFISH by separating quantifications of both nascent and cytoplasmic transcripts (Figure 9). 775 cells were analyzed and the frequency of their cytoplasmic and nascent mRNAs were plotted. In *S. cerevisiae*, nascent RNA and cytoplasmic RNA distribution display a Poisson distribution, characteristic for continuous expression as shown for many genes in *S. cerevisiae*. Therefore, in order to test whether *mdn1*⁺ in *S. pombe* is expressed using the continuous pattern of expression, we tested both nascent and cytoplasmic distributions whether they fit to a Poisson distribution. As shown in Figure 9, they do not fit a Poissonian distribution, not passing the goodness-of-fit χ^2 test. This test tries to fit a Poisson distribution to our samples data.

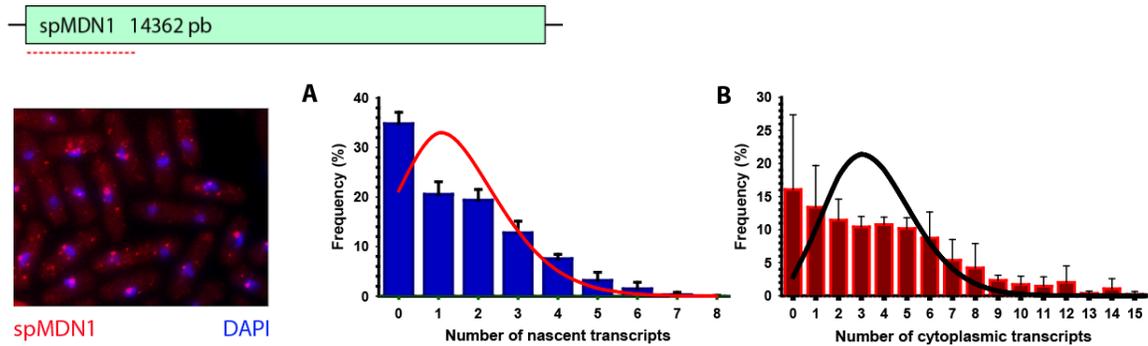


Figure 9. Left: smFISH experiment merged image of *mdn1*⁺ transcripts (Cy3.5) and nuclei (DAPI) in *S. pombe*. A. Frequency distributions of nascent (A) transcripts per allele and cytoplasmic transcripts (B) with superposed fitting of a Poisson distribution. The results were quantified from smFISH performed on *S. pombe* 972 h- strain from three independent experiments (n=775 cells, 1552 Alleles).

As the experimental data did not fit a Poisson distribution, we could rule out the continuous pattern of expression. However, the distributions were still pretty narrow ($\mu=1.55$ and $\sigma=1.56$ for nascent transcripts distribution), very different to distribution observed for strong bursting gene [53, 59]. One characteristic of strong bursting genes is that only a small fraction of cells will show nascent RNAs for a certain gene, whereas continuous expressed genes show most cells having nascent RNAs. We therefore quantified the number of transcription sites present in all the cells. Depending whether cells are haploid or diploid, the continuous pattern would theoretically show a large fraction of cells having one or two transcription sites, while the bursting pattern should show only a few cells with a transcription site, however, those would be more intense. We observed that regardless of the cell-cycle stage, nearly 90% of the 775 cells studied show at least one transcription site (Figure 10). This shows that *mdn1*⁺, although not expressed in a continuous fashion, is not transcribed in a strong bursting pattern.

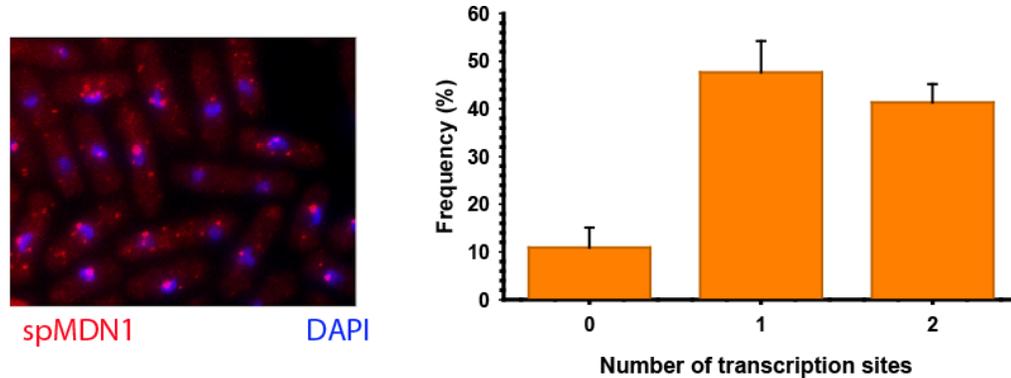


Figure 10. Left: smFISH merged image of nuclei (DAPI) and *mdn1*⁺ transcripts (Cy3.5). Right: Number of transcription sites of *mdn1*⁺ gene in *S. pombe* 972 h- strain from three independent experiments (n=775 cells).

***mdn1*⁺ is expressed by a continuous-like pattern of expression in *S. pombe* in G2 but does not fit a Poissonian distribution**

The cells studied and quantified in the Figure 9 are at different stages of their cell-cycle. We observed an expression pattern that does not fit to a simple Poisson stochastic model but is also not compatible with a strong bursting model. Therefore, we were wondering whether inability to fit to a Poisson distribution might be caused by variation of expression in different cell-cycle stages and that within the individual cell-cycle stages expression would be consistent with a Poisson model.

The yeast *Schizosaccharomyces pombe*, compared to the budding yeast, divides by septation which occurs right in the middle of the dividing cell. When grown in rich media, *S. pombe* 972 h- cells spend about 70% of their lifespan in G2, where the cells grow in length, 10% in M phase, where the nuclei separate and reach the extremity of the dividing cell, and finally 10% in G1/S, as soon as a septum is visible in a dividing cell (Figure 11). When observing cells using microscopy, G1 and S phases are not distinguishable, as they start when septation occurs [90, 91, 117]. Therefore, we can separate cells only in G2, M and G1/S phase and perform mRNA quantification by smFISH at these different cell-cycle stages.

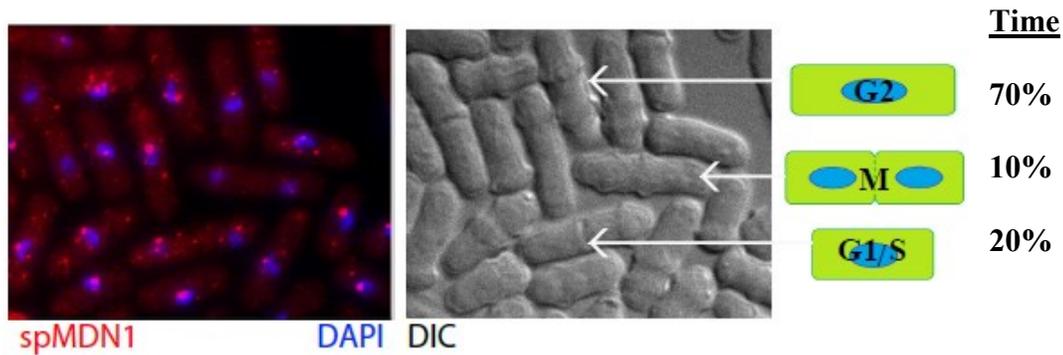


Figure 11. Left: smFISH merged images of nuclei (DAPI) and *mdn1*⁺ (Cy3.5) in *S. pombe* 972 h- cells. Right: Cells from left panel under transmitted light pointed out at different cell-cycle stages as categorized in this project. *S. pombe* 972 h- strain cells spend around 70% of their lifespan in G2 phase, 10% in M phase and 20% in G1 and S phase.

Cells for the analysis shown in Figure 9 were binned into the different cell cycle stages, identifying 535 cells of the 775 total cells found to be in G2 phase. Frequency of nascent and total mRNA levels were plotted and tested for fitting to a Poisson distribution (Figure 12A-B). Neither of the distributions fit a Poisson distribution tested with a χ^2 test. However, similar to what discussed above, distributions are very narrow and do not suggest strong bursting. We then tested whether the nascent transcripts distribution in G2 would fit another mathematical discrete distribution, other than the Poisson. Using the MATLAB function **allfitdist**, we found that our nascent transcripts would fit the best to a negative binomial distribution (Figure 12C), much better than other discrete distributions like a Poisson or a binomial distribution. By definition, a negative binomial distribution is the probability for an event to be successful after a certain number of failures, known as the Bernoulli trial. The only possible outcomes are a success or a failure. When the numbers of successful events become high, the negative binomial distribution approaches a Poisson distribution.

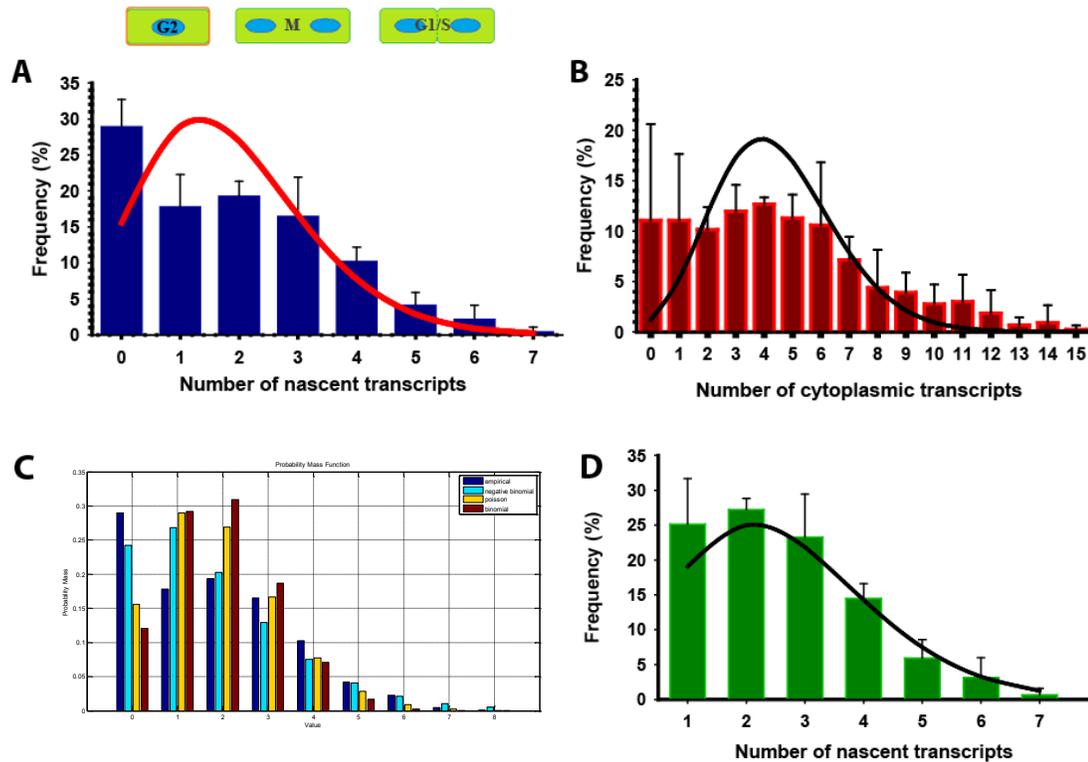


Figure 12. A. Frequency distribution of nascent transcripts in all alleles separately (n=1070 alleles) of cells in G2 phase only with superposed Poisson distribution. Data quantified from three independent smFISH experiments. B. Frequency distribution of cytoplasmic transcripts of *mdn1*⁺ in *S. pombe* 972 h- strain in G2 stage (n=535 cells) with superposed Poisson distribution. C. Result of nascent transcripts of *mdn1*⁺ distribution data tested with MATLAB function `alldistfit` for fitting discrete distribution. D. Frequency distribution of nascent transcripts of cells in G2 phase only without inactive alleles and Poisson distribution fit superposed.

It seemed, from our experimental data and Poisson fitting in Figure 9A, that the number of inactive alleles were shifting the average of the Poisson distribution and were possibly the reason it did not fit a Poisson distribution. Therefore, we tested the same experimental data by removing the inactive alleles. As shown in Figure 12D, when active, the expression of *mdn1*⁺ follows a Poisson distribution in *S. pombe* in G2.

From our experimental data, we found that *mdn1*⁺ is not expressed in a pure stochastic manner during the G2 phase in *S. pombe*. Nevertheless, it does not fit a robust

bursting pattern either. Our data show that *mdn1*⁺ is expressed in a pure stochastic manner once the expression is initiated (Figure 12-D). This suggests that *mdn1*⁺ could be expressed by very small bursts during G2 phase since the distribution shown in Figure 12-D would not have fitted a Poisson distribution if the gene was expressed in a strong bursting pattern.

Following *mdn1*⁺'s expression during M and G1/S phase in *S. pombe*

mdn1⁺ is a constitutively active gene, which means it is expressed at all cell-cycle stages. As its expression did not follow a pure Poisson distribution during G2 phase, as shown previously, we wanted to investigate the transcription mode in the other cell-cycle stages. We expected to find the same results as in G2 phase. As expected, the distributions found in M phase for both the nascent and cytoplasmic transcripts did not fit a Poisson distribution (Figure 13A). Surprisingly, in G1/S the nascent transcripts distribution does fit a Poisson distribution ($p=0.4673$, $\chi^2=2.5444$).

We observed that the amounts of cells that do not transcribe *mdn1*⁺ in M phase are much higher than during G2 phase (Figure 12A). This population lowers as the cells progress in G1/S (Figure 13D). The cytoplasmic population is also narrower during M phase and rises in G1/S phase (Figure 13B, D). This suggests a certain slowdown of expression of *mdn1*⁺ at M phase and a resuming during G1/S phase.

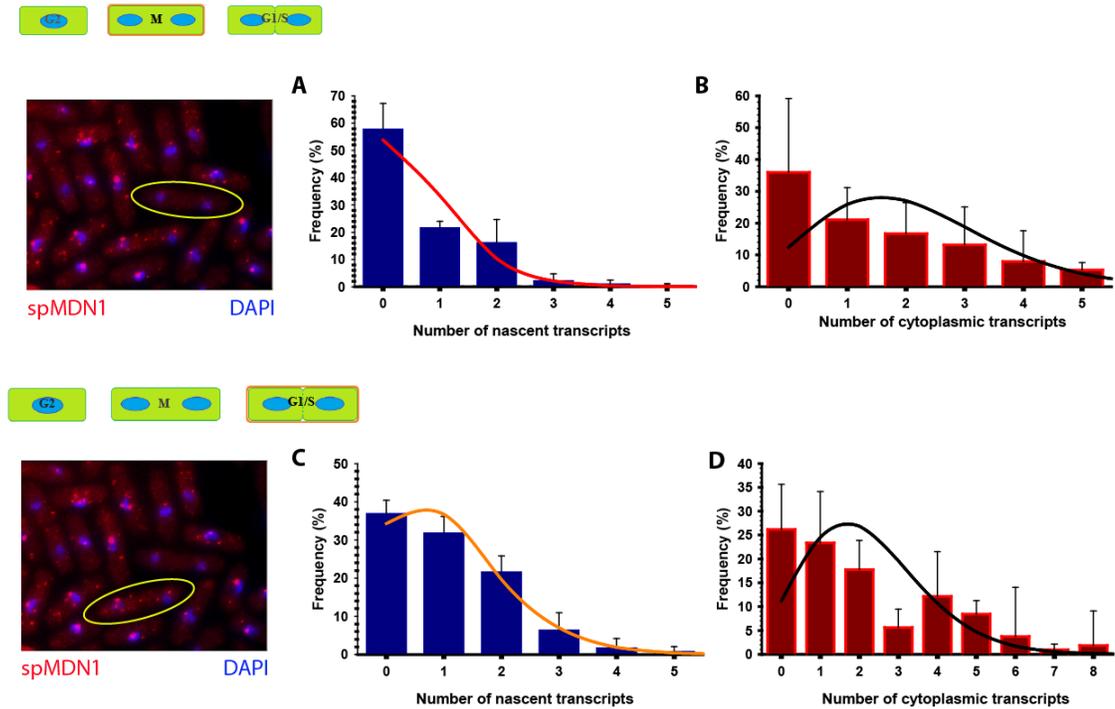


Figure 13. Left panels: smFISH merged images of nuclei (DAPI) and *mdn1*⁺ (Cy3.5) in *S. pombe* 972 h- cells with, circled in yellow, cells in M phase (top) and G1/S phase (bottom). Frequency distributions of nascent transcripts per allele of *mdn1*⁺ in *S. pombe* 972 h- during M phase (A) and G1/S phase (C) with a fit of a Poisson distribution superposed. Frequency distributions of cytoplasmic transcripts of *mdn1*⁺ in *S. pombe* 972 h- strain in M stage (n=114 cells) (B) and G1/S phase (n=100 cells) (D) with a superposed fit for Poisson distribution. All the results shown are quantified from three independent experiments.

We finally compared the distributions for both nascent and cytoplasmic transcripts at all cell-cycle stages, as shown in Figure 14. We observe that for both nascent and cytoplasmic distributions, the G2 cell-phase shows the widest distributions and transcription slows down during M phase and resumes in G1/S phase. Nascent transcripts distributions at all cell-cycle stages are narrow, not corresponding to a robust bursting expression, although not fitting a pure stochastic process during G2 and M phase.

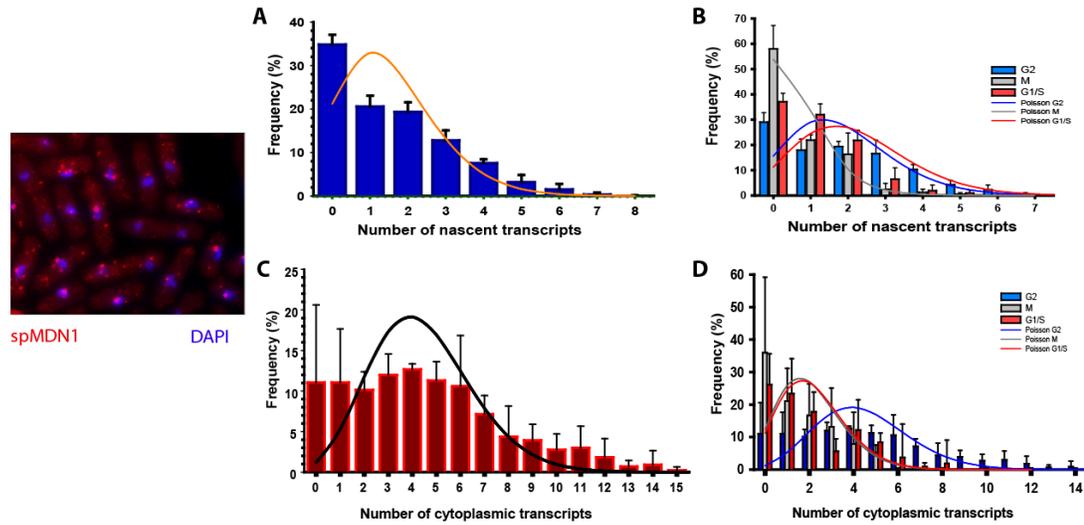


Figure 14. Left: smFISH image of merged nuclei (DAPI) and *mdn1*⁺ (Cy3.5) in *S. pombe* 972 h- cells. Distributions of nascent (A) transcripts and cytoplasmic (C) transcripts of *mdn1*⁺ in *S. pombe* and superposition of a Poisson distribution. Binning and superposition respectively of nascent (B) and cytoplasmic (D) transcripts of *mdn1*⁺ in G2 phase (blue), M phase (grey) and G1/S phase (red).

Fano factor for each phases of the cell-cycle for *mdn1*⁺ in *S. pombe*

We finally calculated the Fano factor, which is a measure of how a distribution is disperse from a Poisson distribution (Table 1). It calculates the variance of an experimental distribution over the mean of the distribution (Equation 1). In that sense, a Poisson distribution will have a Fano factor of 1 which means that the closer the Fano factor is to 1, the closer an experimental distribution is to a Poisson distribution. It has been previously used to define transcription modes and burst size [118, 119]. The Table 1 shows that when all cells are taken together, regardless of cell-cycle stage, the Fano factor of nascent transcripts equals 1.5806. This shows that the burst size is not big enough to be considered as a robust burst, while not being as close to 1 as during G1/S phase.

$$F = \sigma^2 / \mu$$

Eq. 4

Phase	Fano Factor Nascent	Fano Factor Cytoplasmic	Poissonian?
All	1.5806	3.2512	No
G2	1.4986	3.0426	No
M	1.3344	2.6677	No
G1/S	1.0871	2.2696	Yes

Table 1. Calculated Fano factor for each stages of the cell-cycle for *mdn1*⁺ nascent and cytoplasmic transcripts distributions and results of fitting to a Poisson distribution by a chi-square test on *S. pombe* 972 h- strain.

PRP8/spp42⁺ displays a continuous expression in *S. pombe*, following a Poisson process

Observing a different expression mode for *mdn1*⁺ in *S. pombe* compare to *S. cerevisiae*, we wondered if this is a common phenomenon in *S. pombe*, and tested for the expression of a second essential gene, encoding for the splicing factor Prp8. Named *spp42* in *S. pombe*, the protein has a central role in the catalytic core of the spliceosome [120]. Equal to *mdn1*⁺, *PRP8/spp42*⁺ expression displays a continuous pattern in *S. cerevisiae* [53].

Cytoplasmic and nascent transcript distributions were determined using smFISH as described above for the *mdn1*⁺ gene. Different from *mdn1*⁺, we have found that nascent transcripts distribution of *PRP8/spp42*⁺ does fit a Poisson distribution, while cytoplasmic transcripts distribution does not (Figure 15).

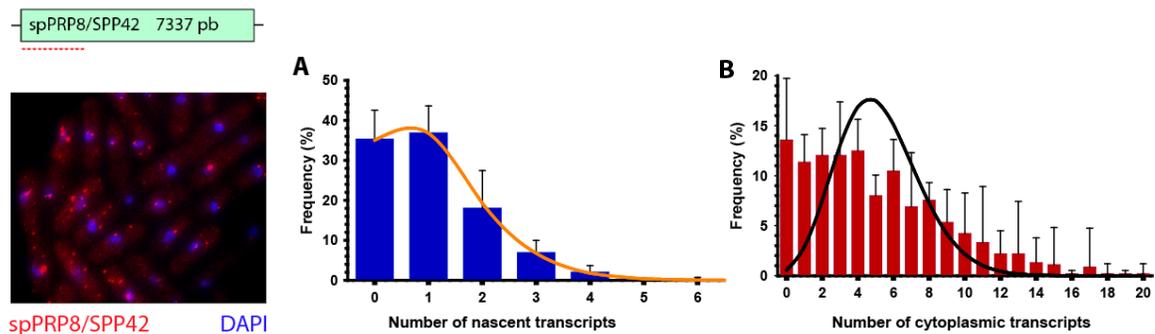


Figure 15. Left: smFISH experiment merged image of *PRP8/spp42*⁺ transcripts (Cy3.5) and nuclei (DAPI) in *S. pombe*. A. Frequency distributions of nascent (A) transcripts per allele and cytoplasmic transcripts (B) with superposed fitting of a Poisson distribution. The results were quantified from smFISH performed on *S. pombe* 972 h- strain from three independent experiments (n=524 cells).

We also analyzed the number of transcription sites of the *prp8* gene in all cells following the same logic as done for *mdn1*⁺ (Figure 16). In *S. pombe*, more than to 85% of the population studied has at least one transcription site for *PRP8/spp42*⁺. Just as in *mdn1*⁺, most cells are transcribing *PRP8/spp42*⁺, supporting a continuous pattern of expression.

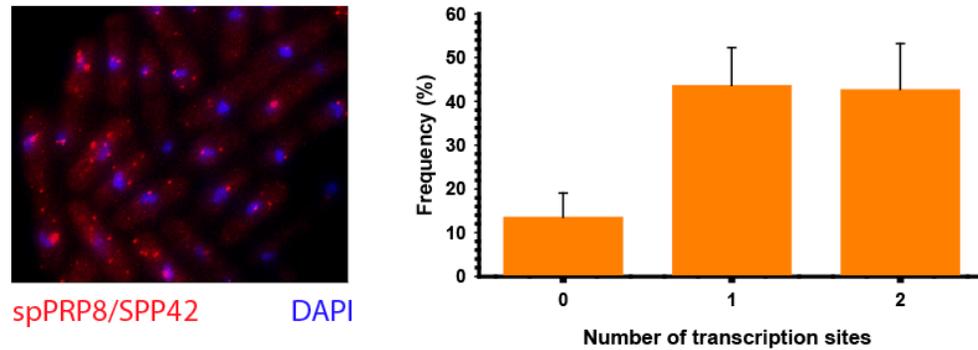


Figure 16. Left: smFISH merged image of *PRP8/spp42*⁺ transcripts (Cy3.5) and nuclei (DAPI) in *S. pombe* 972 h- strain. Right: Number of transcription sites of *PRP8/spp42*⁺ gene in *S. pombe* 972 h- strain from three independent experiments (n=524 cells).

***PRP8/spp42*⁺ is expressed in a continuous pattern of expression in *S. pombe* at all stages**

As we have assessed that *PRP8/spp42*⁺ is expressed in a continuous pattern, we wanted to know if at all cell-cycle stages it fit a Poisson process. We expected to find *PRP8/spp42*⁺ expressed in a stochastic manner in G2, M and G1/S.

We started by quantifying the nascent and cytoplasmic transcripts of *PRP8/spp42*⁺ in G2 phase, as it is the phase *S. pombe* 972 h- strain spends 70% of its lifespan. We tested the distributions found for a Poisson distribution by a χ^2 test. Nascent transcripts frequency distribution fits a random Poisson distribution, while cytoplasmic does not (Figure 17).

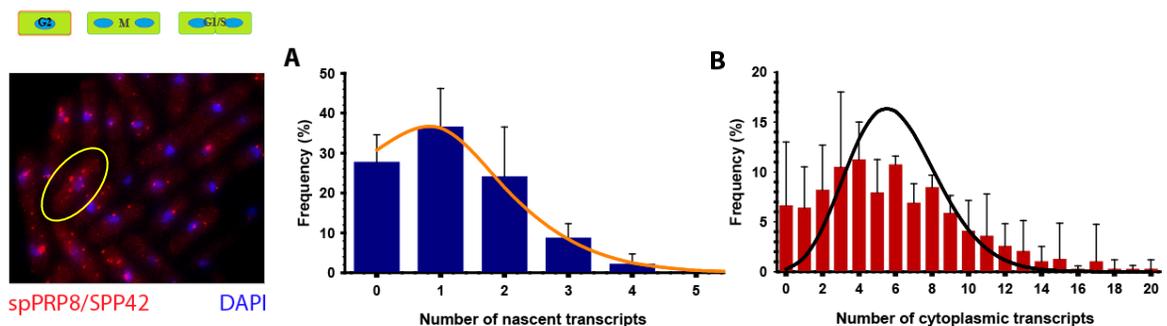


Figure 17. Left: merged smFISH with *PRP8/spp42*⁺ mRNA (Cy 3.5) and nuclei (DAPI). A cell in G2 phase is circled in yellow. A. Frequency distribution of nascent transcripts per allele of *PRP8/spp42*⁺ in *S. pombe* 972 h- strain in G2 phase and a superposed fitting of a Poisson distribution on the nascent transcripts per allele distribution in G2 ($p=0.8205$, $\chi^2_{\text{stat}}=1.5347$) B. Frequency distribution of cytoplasmic transcripts of *PRP8/spp42*⁺ in *S. pombe* 972 h- strain in G2 stage (n= 392 cells) with superposed fitting of a Poisson distribution.

We finally studied the pattern of expression of *PRP8/spp42*⁺ in *S. pombe* in M and G1/S phase. The frequency distributions of nascent transcripts per allele in both studied phases did fit a Poisson distribution (Figure 18 A, C). As previously shown in *mdn1*⁺, the number of cells containing no *PRP8/spp42*⁺ transcription site increases during M phase and lowers in G1/S. The cytoplasmic distributions reflect the same thing as the number of cells containing no *PRP8/spp42*⁺ transcript is higher in M phase than in G1/S phase (Figure 18 B, D).

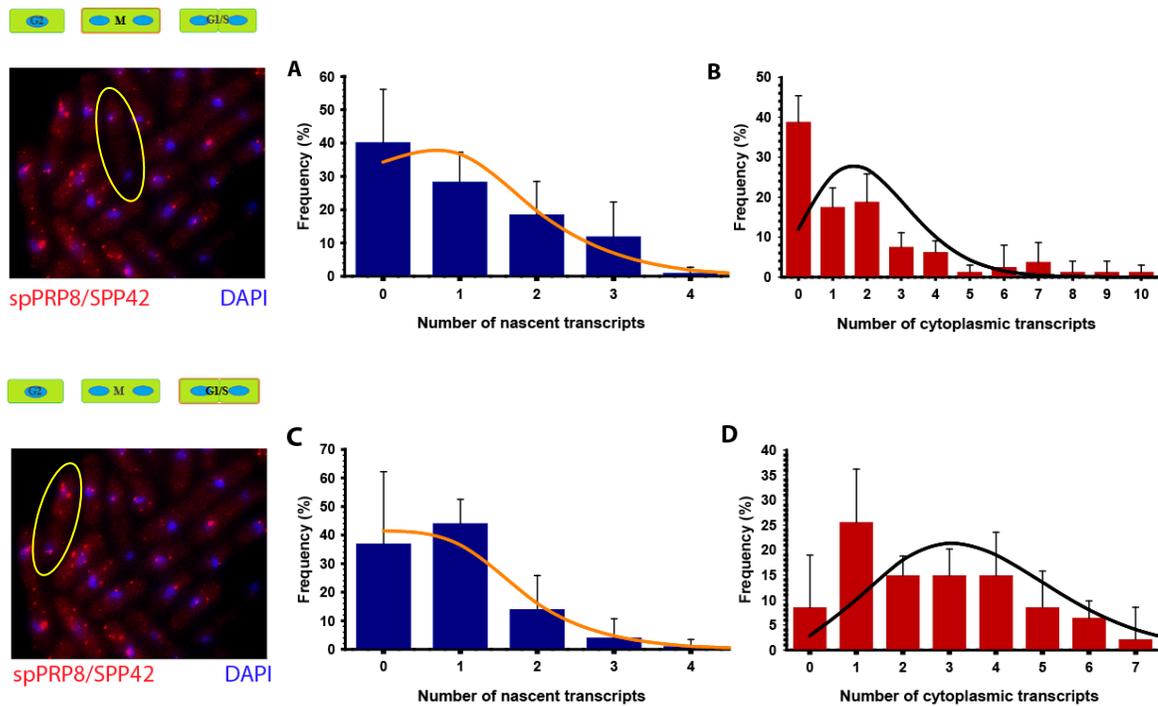


Figure 18. Left: merged smFISH with *PRP8/spp42*⁺ mRNA (Cy 3.5) and nuclei (DAPI) with yellow circled cell in M phase (top) and G1/S phase (bottom). Frequency distributions of nascent transcripts per allele of *PRP8/spp42*⁺ in *S. pombe* 972 h- strain in M phase (A) and in G1/S phase (C) and a superposed fitting of a Poisson distribution on the nascent transcripts for M phase ($p = 0.1960$ $\chi^2_{stat} = 3.2591$) and G1/S phase ($p = 0.2947$ $\chi^2_{stat} = 2.4433$) Frequency distributions of cytoplasmic transcripts of *PRP8/spp42*⁺ in *S. pombe* 972 h- strain in M phase (B) ($n = 82$ cells) and G1/S phase (D) ($n = 50$ cells) with superposed fittings of a Poisson distribution.

Following $mdn1^+$ and $PRP8/spp42^+$ during M phase in *S. pombe*

The genes $mdn1^+$ and $PRP8/spp42^+$ are two essential and constitutively expressed genes in all organisms. Their expression should then not vary much during the cell-cycle, as they are not influenced by the latter [62, 121, 122]. However, we observed for both genes a diminishment of expression for both genes during M phase (Figure 14 and 18).

To understand how their expression varies in a time-course manner during M phase, we calculated the coefficient of the distance between the two nuclei (d) over the total length of the cell (e). As cells progress in M phase, the coefficient d/e becomes closer to 1. We binned the different measures into four groups and compared the number of average nascent transcripts found at each of the groups for both $mdn1^+$ and $PRP8/spp42^+$ in *S. pombe* (Figure 19).

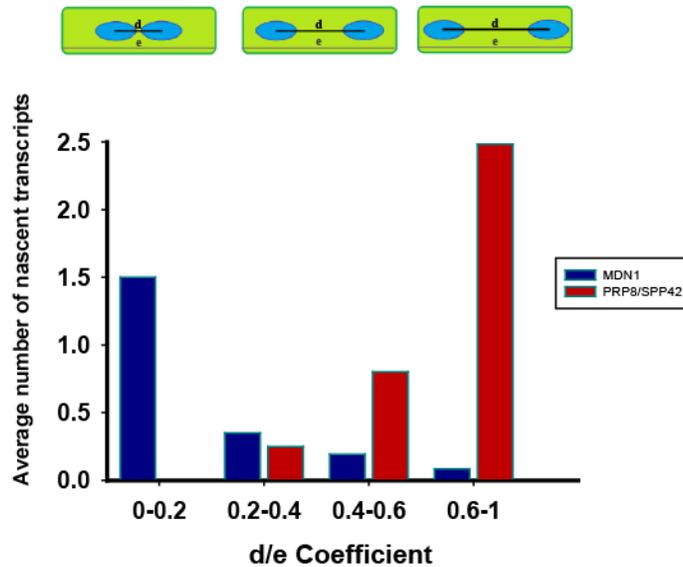


Figure 19. Average nascent transcripts for $mdn1^+$ (A) and $PRP8/spp42^+$ (B) in *S. pombe* during M phase at different d/e coefficients from three independent experiments. For $mdn1^+$ and $PRP8/spp42^+$ respectively, 2% and 5% of cells were found between 0 and 0.2, 13% and 10% between 0.2 and 0.4, 52% and 56% between 0.4 and 0.6, and 33% and 29% between 0.6 and 1.

We finally compared all nascent and cytoplasmic transcripts distributions of *PRP8/spp42*⁺ at the different cell-cycle stages. Figure 20 shows that the nascent and cytoplasmic transcripts distributions get wider during G2 phase and gets narrower during M. In fact, many cells in M show no cytoplasmic transcripts at all. This follows our observation in which during M phase, the expression of *PRP8/spp42*⁺ is shut down, as shown in Figure 19.

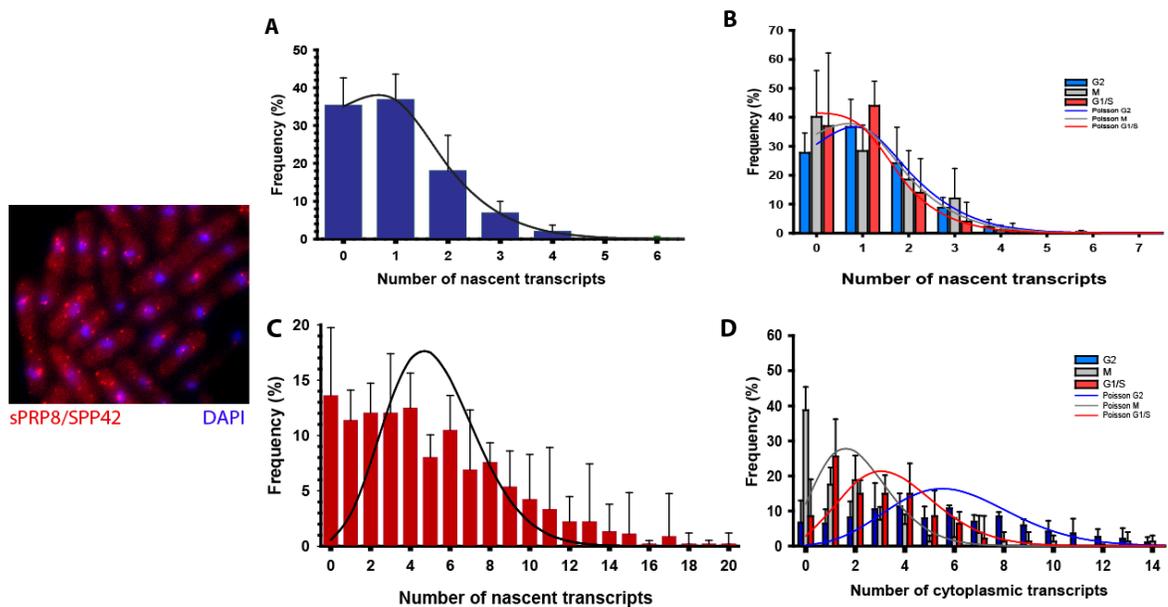


Figure 20. Left: smFISH image of merged nuclei (DAPI) and *PRP8/spp42*⁺ (Cy3.5) in *S. pombe* 972 h- cells. Distributions of nascent (A) transcripts and cytoplasmic (C) transcripts of *PRP8/spp42*⁺ in *S. pombe* and superposition of a Poisson distribution. Binning and superposition respectively of nascent (B) and cytoplasmic (D) transcripts of *PRP8/spp42*⁺ in G2 phase (blue), M phase (grey) and G1/S phase (red).

Fano factor for each phases of the cell-cycle for *PRP8/spp42*⁺ in *S. pombe*

Finally, we calculated the Fano factor at each cell-cycle stage for the nascent *PRP8/spp42*⁺ transcripts distributions (Table 2). We observe that the Fano factor at each phase is very close to 1, especially when all cells grouped together.

Phase	Fano Factor Nascent	Fano Factor Cytoplasmic	Poissonian?
All	1.0493	3.9788	Yes
G2	0.9994	3.559	Yes
M	1.0993	4.0307	Yes
G1/S	0.8558	3.0057	Yes

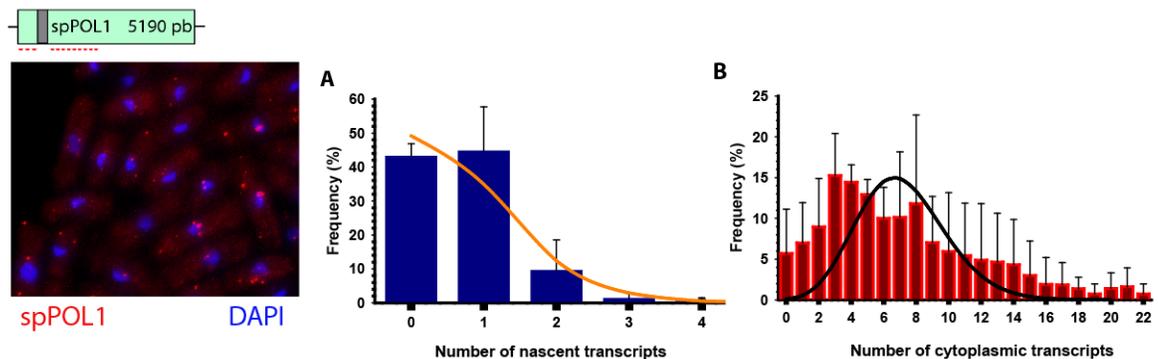
Table 2. Calculated Fano factor for each stages of the cell-cycle for *PRP8/spp42*⁺ nascent transcripts distributions and results of fitting to a Poisson distribution by a chi-square test on *S. pombe* 972 h- strain.

pol1⁺ and *cdc13*⁺'s expression do not fit a Poisson distribution

In theory, both *mdn1*⁺ and *PRP8/spp42*⁺ are genes that should be constitutively active genes and expressed at all cell-cycle stages. We have seen that there seems to be a cell-cycle component to their expression. We observed that although *mdn1*⁺ has a pattern of expression approaching a stochastic process, *prp8* is expressed in a Poisson manner.

Many genes are known to be transcribed in a cell-cycle specific manner. In *S. cerevisiae*, it was shown that *pol1*⁺, the gene coding for DNA polymerase, is transcribed in a cell-cycle specific manner, however during the time it is active, its transcription is stochastic [53]. In order to know if cell-cycle regulated genes behave the same in *S. pombe*, we studied two cell-cycle regulated genes, *cdc13*⁺ and *pol1*⁺. First we determined the expression mode of *cdc13*⁺, encoding for the *cdc13*⁺ protein, also known as Cyclin B in humans. It is an essential protein with a central role in the transition from G2 to M phase of the cell-cycle [123]. This gene's transcription kinetics was studied in *S. cerevisiae*. It has been shown that *cdc13*⁺/CLB2 transcripts accumulate progressively during and rapidly degraded upon M phase [124, 125].

We first plotted the frequency of nascent transcripts found per allele in all cells as well as cytoplasmic transcripts of *pol1*⁺ and *cdc13*⁺, regardless of their cell-cycle stage (Figure 21). As expected, none of the distributions fit a Poisson distribution when tested by a χ^2 test. Interestingly, few cells show no cytoplasmic RNAs, indicating either transcription is not shut down entirely or mRNA do not get degraded efficiently.



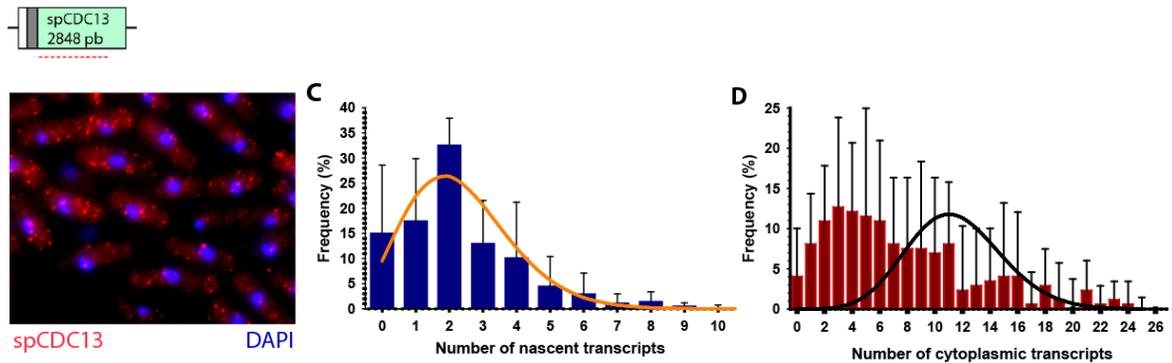
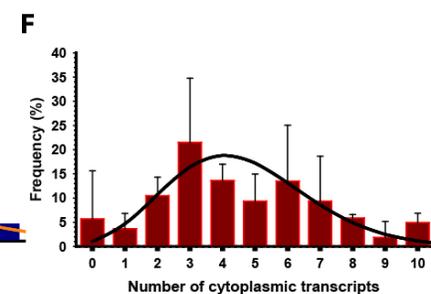
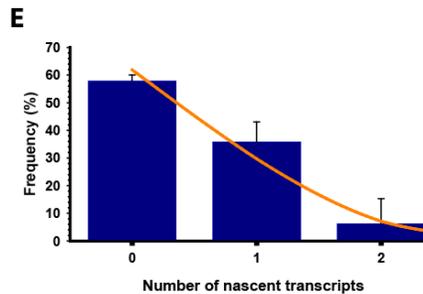
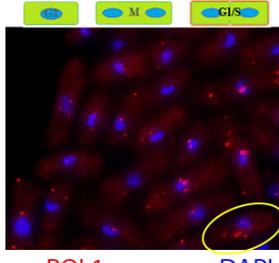
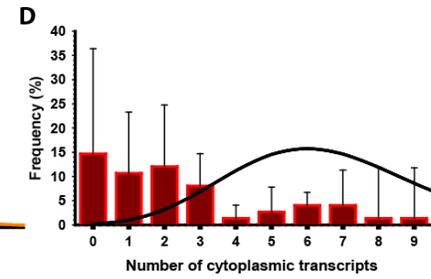
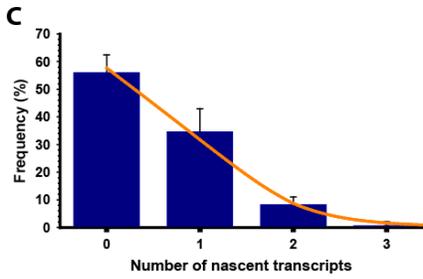
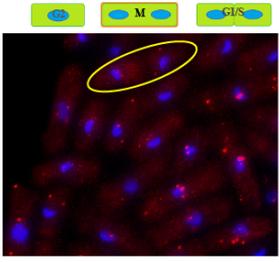
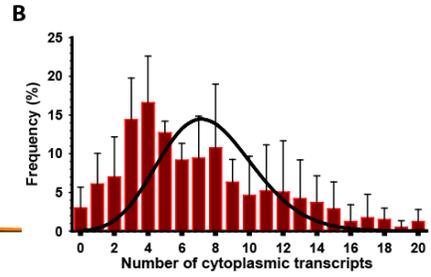
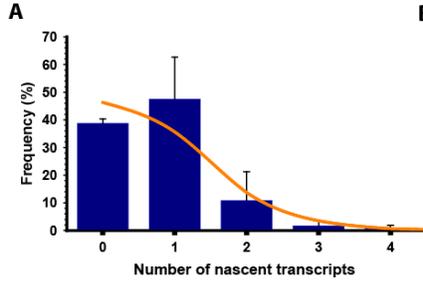
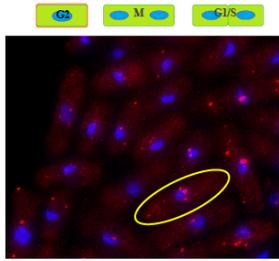


Figure 21. Left panels: smFISH merged image of nuclear signal (DAPI) and *pol1*⁺ (top) or *cdc13*⁺ mRNAs (Cy3.5) with respective gene length's and structure representation. Frequency distributions of nascent *pol1*⁺ (A) and *cdc13*⁺ (C) transcripts per allele in *S. pombe* 972 h- and fitting of a Poisson distribution for each. Frequency of cytoplasmic transcripts of *pol1*⁺ (B) (n= 669 cells) and *cdc13*⁺ (D) (n= 656 cells) and fitting of a Poisson distribution for each. The results were quantified from smFISH performed on *S. pombe* 972 h- strain from three independent experiments.

Expression of *pol1*⁺ and *cdc13*⁺ is at their highest during G2 phase

Following the same logic as for the previously studied genes, we binned the cells shown in Figure 21 separately in three different stages as G2 (Figure 22 A, B, G, H), M (Figure 22 C, D, I, J) and G1/S (Figure 22 E, F, K, L). Only one distribution fit a Poisson process: nascent *pol1*⁺ transcripts during M phase. All the other distributions did not fit when tested with a χ^2 test.



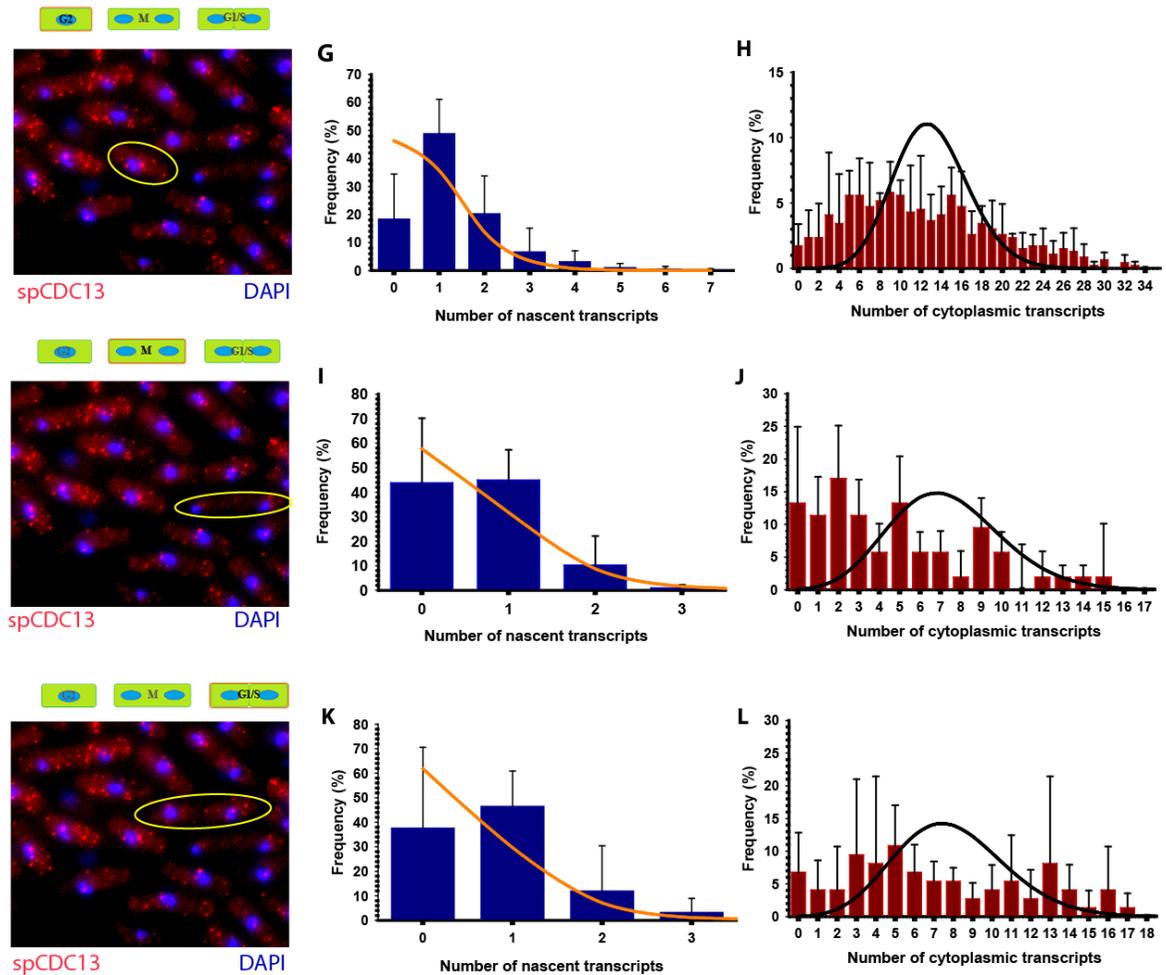


Figure 22. Left panels: smFISH merged image of nuclei (DAPI) and $poll^+$ or $cdc13^+$ transcripts in *S. pombe* cells. Cells found at different cell-cycle stages are circled in yellow for each gene. Frequency of nascent transcripts of $poll^+$ at phase G2 (A), M (C), and G1/S (E); $cdc13^+$ transcripts at phase G2 (G), M (I), and G1/S (K) with superposed fitting of Poisson distribution. Frequency of cytoplasmic transcripts of $poll^+$ at phase G2 (B) (n=498 cells), M (D) (n=92 cells), and G1/S (F) (n=84 cells); $cdc13^+$ at phase G2 (G) (n=472 cells), M (I) (n=104 cells), and G1/S (K) (n=148 cells) with superposed fitting of Poisson distribution. All the results shown are quantified from three independent experiments.

As it was suggested by microarray experiments, the peak of transcription of both $poll^+$ and $cdc13^+$ is during M phase in *S. pombe*, we aimed to assess this using smFISH [121, 122]. We calculated that in average, respectively for $poll^+$ and $cdc13^+$: in G2 0.77 and 1.34 nascent transcripts per allele are produced, while 0.55 and 0.73 nascent transcripts

per allele are produced in M phase and 0.48 and 0.8 transcripts per allele are produced in average during G1/S phase.

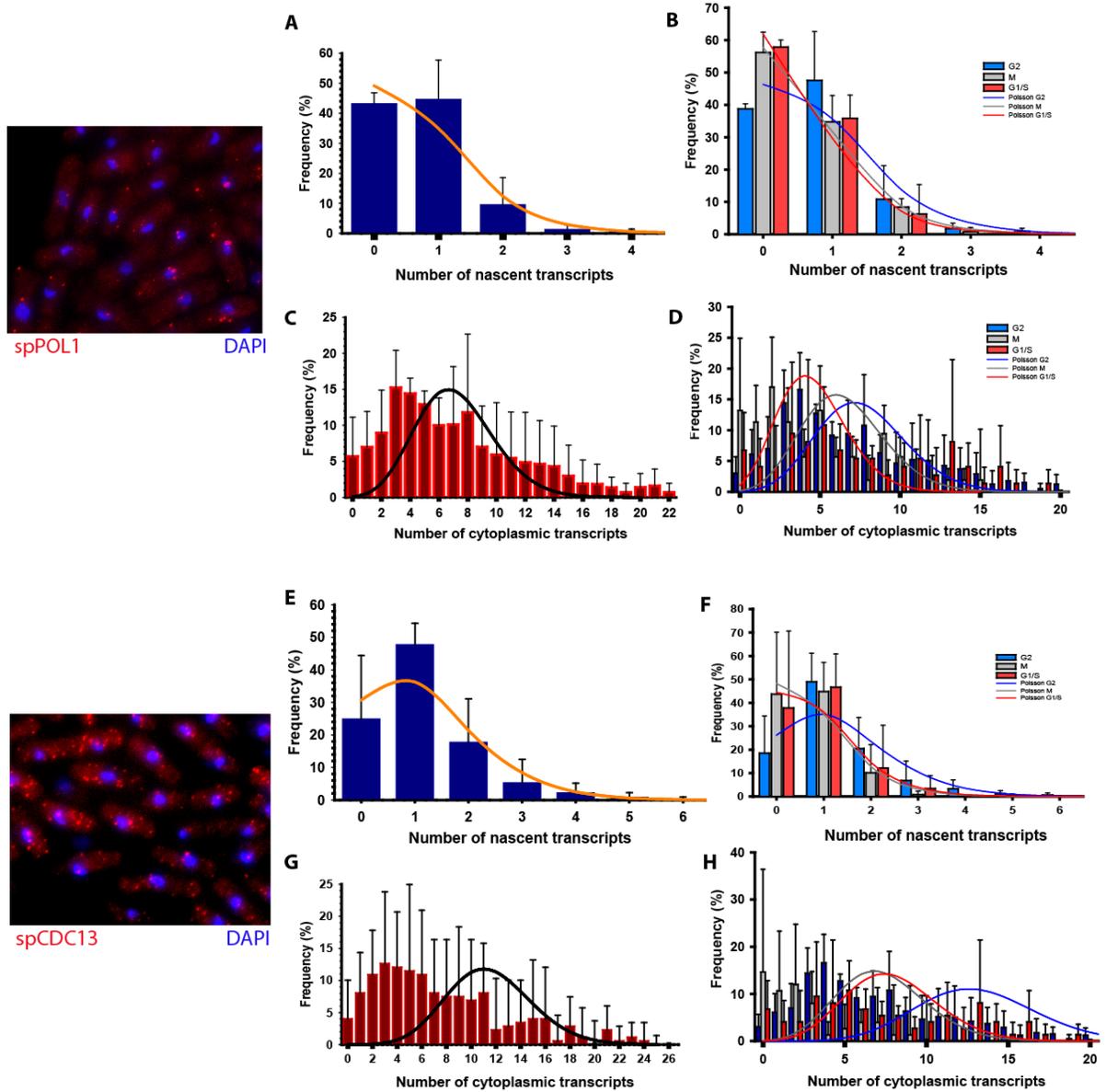


Figure 23. Left panels: smFISH merged image of nuclei (DAPI) and *poll*⁺ or *cdc13*⁺ transcripts in *S. pombe* cells. Distributions of nascent transcripts of *poll*⁺ (A) and *cdc13*⁺ (E) and cytoplasmic transcripts of *poll*⁺ (C) and *cdc13*⁺ (G) in *S. pombe* and superpositions of a Poisson distribution. Binning and superposition respectively of *poll*⁺ (B) and *cdc13*⁺ (F) nascent as well as *poll*⁺ (D) and *cdc13*⁺ (H) cytoplasmic transcripts in G2 phase (blue), M phase (grey) and G1/S phase (red).

When comparing the different distributions, in Figure 23, we can appreciate that in G2 the number of inactive alleles are the lowest and the cytoplasmic transcripts distribution is wider at that stage for both genes. We observe that the nascent transcripts distributions look very similar, although not all fitting a Poisson process.

Fano factor for each phases of the cell-cycle for $poll^+$ and $cdc13^+$ in *S. pombe*

Finally, we calculated the Fano factor for the nascent and cytoplasmic transcripts of $poll^+$ and $cdc13^+$ at each stage of the cell-cycle. We can observe that the Fano factor is never greater than 1, but the closest to 1 occurs in G2 phase, where we observe that the expression initiation of $poll^+$ and $cdc13^+$ is at its peak.

Gene	Phase	Fano Factor Nascent	Fano Factor Cytoplasmic	Poissonian?
$poll^+$	All	0.8226	4.1972	No
	G2	0.8147	4.0678	No
	M	0.8119	5.1682	Yes
	G1/S	0.7281	2.4931	No
$cdc13^+$	All	0.9668	6.9113	No
	G2	0.938	6.6363	No
	M	0.7862	6.2158	No
	G1/S	0.7442	4.0574	No

Table 3. Calculated Fano factor for each stages of the cell-cycle for $poll^+$ and $cdc13^+$ nascent transcripts distributions and results of fitting to a Poisson distribution by a chi-square test on *S. pombe* 972 h- strain.

Studying gene expression patterns in Human cell-lines by smFISH

As results were obtained in the yeast *S. pombe*, we wanted to pursue our study of understanding how gene expression patterns have evolved with evolution in other eukaryotic organisms (see Rationale). However, establishing smFISH and data analysis in *S. pombe* demanded more time than initially expected. As human cell-lines are very accessible and constitute the ultimate studied model of our project, we have decided to study the expression of MDN1, PRPF8, POLA1 and CCNB1, all homologs of the previously presented genes.

We have designed and labelled probes for each of the genes, spanning on the exons while avoiding the exon-exon junctions, to observe and eventually quantify the pre-mRNAs (nascent mRNAs). We have tested all the probes sets but here we show only smFISH images for MDN1 (Figure 24). We implemented smFISH in the human cell-lines HEK293T, U2OS and IMR90. HEK293T cells are embryonic kidney cancerous cells expressing the SV40 T-antigen. They are widely used as they are highly transfectable and their metabolism is known to be increased compared to other cancerous cell-lines. They are adherent, making them a good cell-line for imaging. U2OS are derived from human bone osteosarcoma and are also suited for microscopy as they adhere well. However, the major inconvenient with cancerous cell-lines is their polyploidy. Both cancerous cell-lines used are nearly-triploid, explained by the fact that they are usually tetraploid but lose some chromosomes and the latter rearrange [126]. Therefore, the HEK293T and U2OS cell-lines do not represent precisely physiological conditions. Thus, we used given cells from a primary cell-line, IMR90, derived from normal lung fibroblasts.

We used two sets of probes: one hybridizing the 5' region of a given mRNA, and the other targeting the 3' region of the same mRNA. This allows us to precisely observe and quantify transcription sites for the genes we study (Figure 24).

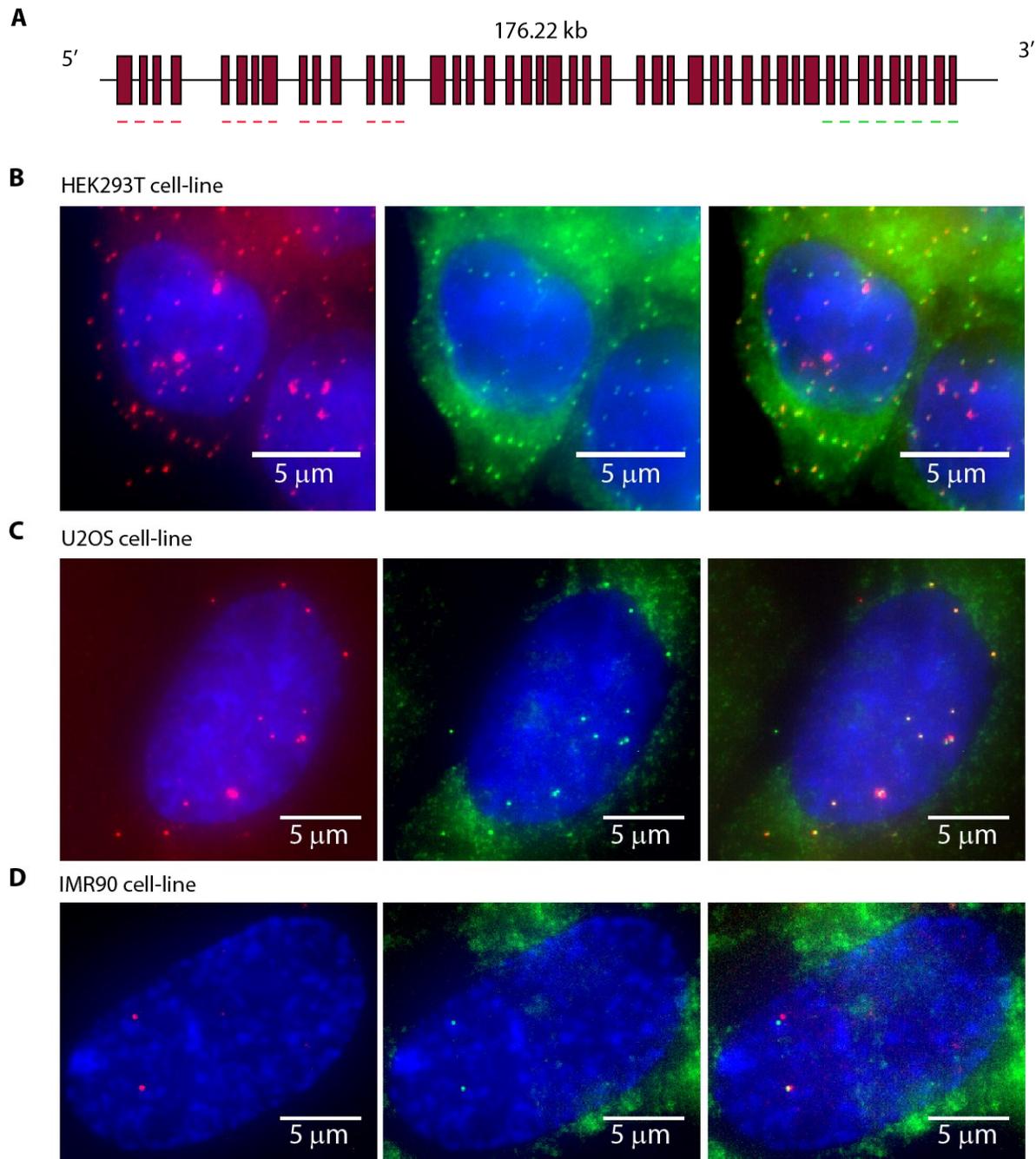


Figure 24. (A) Representation of MDN1 gene with exons (dark red rectangles) and introns and disposition of 5' probe set (red dotted line) and 3' probe set (green dotted line). (B-D) smFISH experiments on Human cell-lines for *mdn1*⁺ gene and nuclear signal by DAPI. (B Left to right) 5' exons in Cy3, 3' exons in Cy3.5 and merged signals for HEK293T cells. (C Left to right) 5' exons in Cy3, 3' exons in Cy3.5 and merged signals for U2OS cells. (D Left to right) 5' exons in Cy3, 3' exons in Cy3.5 and merged signals for IMR90 cells.

In order to quantify mRNAs in human cell-lines, we tried using the same scripts as for *S. pombe*. However, CellProfiler segments cells and nuclei in two dimensions. As the cells we image form a monolayer and are not isolated, a two-dimensional segmentation would be too imprecise. Therefore, we used Imaris (Bitplane), a software that allows segmentation of cells and nuclei in three dimensions.

Using segmentation in three dimensions was not compatible with Localize, as it is being done in two dimensions. Thus, we used an equivalent of Localize that detects spots in 3D using the z-stack of a field. We used FISH-Quant, a MATLAB based script developed by Florian Mueller. As the detected spots' information needed to be merged with the three dimensional segmentation by Imaris, Emmanuel Noutahi developed scripts allowing the importation of the spots into Imaris in order to proceed to final quantifications. Both 5' and 3' spot sets are imported in Imaris, using the Imaris XT platform, and the intensity of a single spot was calculated as the mean of cytoplasmic transcripts' intensity (Figure 25).

To determine the localization and quantification of nascent transcripts at transcription sites, we considered three scenarios. The first, a single 5' spot without a 3' spot very close, suggesting a single nascent transcript, must have at least a 1.5 higher intensity than a single transcript to be considered as a transcription site. The second has multiple 5' spots in the same vicinity with or without a 3' spot. We have developed a tool that allows to manually determine a radius around each 5' spots within which, if more than one 5' spot is found, is considered a transcription site. Finally, a 5' spot can be colocalized to a 3' spot but can have at least a 1.5 higher intensity than a single transcript. We consider the latter as a transcription site.

The limiting factor with human cell-line quantification, as described above, is the low amount of cells present per field, as cells are very large.

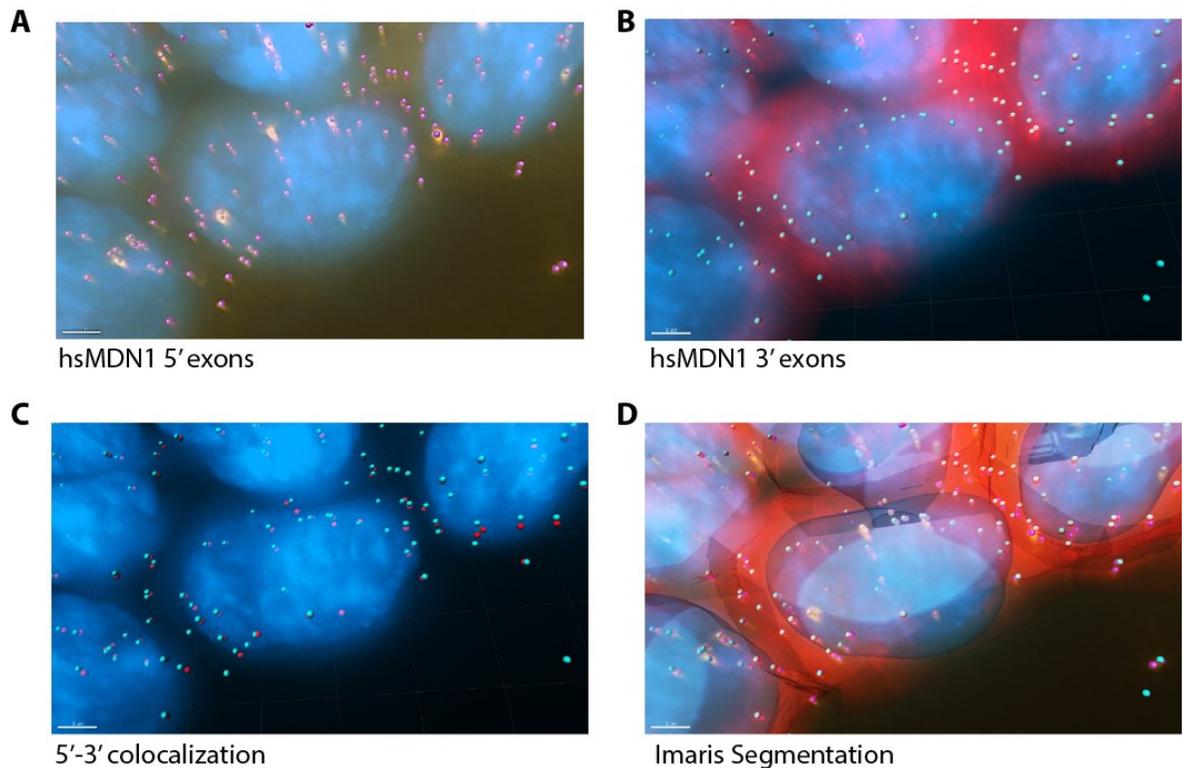


Figure 25. 3D quantification of transcripts in Human cell-lines. A-B. Data of detected 5' exonic and 3' exonic spots, obtained in HEK293T cells, from FISH-Quant were imported into Imaris. The images are a superposition of localized spots in 3D and original Cy3 signal of 5' exonic (A) or Cy3.5 3' exonic (B) MDN1 transcripts and DAPI C. Original DAPI signal and colocalized 5' and 3' spots are then considered, by ARNQuant in Imaris developed by Emmanuel Noutahi, as a single and complete transcript. D. Original DAPI, Cy3 for 5' exonic and Cy3.5 for 3' exonic MDN1 transcripts merged with segmented cells and nuclei by Imaris.

We have observed that transcription sites for MDN1 are more intense in HEK293T cells compared to U2OS cells and to IMR90 cells (Figure 24). Preliminary quantifications suggest that the transcription of MDN1 in HEK293T cells is continuous but may not fit a pure Poisson distribution. However, we do not have a large enough dataset to be able to make a quantitative statement. Further analysis need to be made to determine precisely the pattern of expression of the four genes of interest. We also observe that most transcription sites have a 3' signal.

D. Discussion

Le but premier de ce projet est de comprendre si les patrons d'expression de gène sont gène ou organisme spécifiques. Nous avons donc voulu déterminer l'expression de quelques gènes présents en copie unique par génome dans différents organismes représentant différents stades évolutifs des organismes eucaryotes. Pendant la durée de ma maîtrise, je n'ai obtenu que des résultats pour la levure *S. pombe* et quelques résultats préliminaires pour des lignées cellulaires humaines, quoi que le but de ce projet soit d'obtenir des résultats pour tous les organismes mentionnés plus haut (voir section Évolution de la complexité des génomes).

Les gènes *mdn1*⁺, *PRP8/spp42*⁺, *poll*⁺ et *cdc13*⁺ ont été étudié dans la levure *S. pombe* après que dans la levure *S. cerevisiae* ces gènes aient été démontrés comme étant exprimés de manière continue [53, 124, 125]. L'expression de ces gènes a été détectée et quantifiée grâce à la technique de smFISH, qui a dû être d'abord optimisée et établie dans la levure *S. pombe*.

Établissement de la technique de smFISH dans *S. pombe*

La technique de smFISH a été préétablie dans plusieurs organismes, entre autres dans la levure *S. cerevisiae* et dans des lignées cellulaires humaines [95, 107]. Par contre, cette technique n'avait pas été utilisée dans l'organisme *S. pombe* avant que cette étude ne la mette en place et, entre temps, nous avons partagé le protocole avec d'autres laboratoires qui ont publié des résultats [127]. Tel que décrites dans la section Résultats, de nombreuses modifications ont dû être faites au protocole de smFISH dans *S. cerevisiae* afin d'obtenir des images de grande qualité permettant la quantification d'ARNm dans des cellules uniques. La Figure 7 représente des images obtenues suite à l'utilisation du smFISH où les

gènes *mdn1*⁺, *PRP8/spp42*⁺, *pol1*⁺ et *cdc13*⁺ ont été ciblés dans les cellules de *S. pombe*. Grâce à cette technique, il nous a été possible de quantifier les différents gènes en question afin de déterminer leur patron d'expression dans la levure à fission. Pour ce faire, nous avons utilisé plusieurs outils tels que CellProfiler, un logiciel permettant la segmentation des cellules et des noyaux dans une image de microscopie, le script Localize développé par Daniel Larson permettant de détecter les signaux à partir d'une image de smFISH et enfin, le script ARNQuant développé par Emmanuel Noutahi permettant le lien entre les deux scripts précédemment présentés (Figure 8). Ensemble, nous avons créé des outils permettant de quantifier de manière efficace l'expression des ARNm dans *S. pombe*.

Le gène *mdn1*⁺ n'est pas exprimé de manière continue ni par *burst* robuste dans la levure *S.pombe*

Nous avons débuté notre étude avec le gène *mdn1*⁺ en quantifiant séparément les transcrits naissants et les transcrits cytoplasmiques. Nous avons ensuite transposé le nombre de transcrits quantifiés selon la fréquence dont ils ont été retrouvés (Figure 9). Il a été démontré précédemment dans la levure *S. cerevisiae* que la distribution de transcrits naissants possède plus de pouvoir à prédire un patron d'expression d'un gène plutôt que la simple observation des ARNm cytoplasmiques [53]. Les transcrits cytoplasmiques ne donnent pas la même information sur l'initiation de la transcription, puisque l'observation de ces derniers dépend de leur temps de demi-vie. Par contre, ils peuvent renseigner sur le nombre de transcrits ayant été produits pendant leur temps de vie au moment où les images de smFISH ont été prises. Certains processus tels que la dégradation des transcrits influence le nombre d'ARNm cytoplasmiques alors qu'elle n'affecte pas les ARNm naissants. Nous attribuons donc plus d'importance à la correspondance entre les distributions de transcrits naissants à une distribution de Poisson que les distributions de transcrits cytoplasmiques. Les distributions présentées dans cette section ont toutes été testées par le test statistique de χ^2 afin de déterminer si celles-ci correspondent à une distribution de Poisson, soit tout à fait aléatoire.

Ainsi, les distributions présentées à la Figure 9 représentent les transcrits naissants (Figure 9A) et les transcrits cytoplasmiques (Figure 9B) du gène *mdn1*⁺ dans la levure *S. pombe* 972 h- indépendamment du cycle cellulaire dans lequel les cellules se trouvent. Ces deux distributions ne correspondent pas à une distribution de Poisson, indiquant que du moins l'initiation de la transcription de *mdn1*⁺ n'est pas purement stochastique.

De plus, nous avons remarqué la grande variabilité présente dans le nombre des transcrits cytoplasmiques (Figure 9B). Il est possible d'observer que certaines cellules ont jusqu'à 15 transcrits cytoplasmiques. Il a été calculé par des expériences de micropuces que le temps de demi-vie de l'ARNm *mdn1*⁺ dans *S. pombe* est de 43 minutes [128]. Cela signifie que lorsque les cellules ont été fixées, les ARNm cytoplasmiques représentent les initiations entamées au cours d'une très longue période. Ceci étant dit, nous n'observons pas de distribution bimodale pour la distribution des ARNm cytoplasmiques, ni pour la distribution des transcrits naissants.

Nous avons donc voulu savoir combien d'allèles au total transcrivait *mdn1*⁺. La Figure 10 nous renseigne surtout sur le nombre d'allèle où la transcription du gène *mdn1*⁺ est activée dans la population de cellules étudiées. Ainsi, près de 90% des cellules expriment *mdn1*⁺, ce qui suggérerait un mode d'expression en continu, alors que les distributions de transcrits naissants ne concordent pas avec une distribution de Poisson. Il est donc difficile de trancher entre une expression continue du gène *mdn1*⁺ ou en *burst* chez l'organisme *S. pombe*.

Selon cette même étude utilisant des micropuces, il y aurait en moyenne 4.5 ARNm de *mdn1*⁺ par cellule de *S. pombe* [128]. De manière plutôt concordante, nous avons calculé grâce au smFISH qu'il y en aurait en moyenne 3.5. Nous pouvons donc valider notre technique smFISH et la quantification des transcrits que nous avons élaborée dans les cellules de *S. pombe* comme étant précises.

L'expression du gène *mdn1*⁺ ne correspond pas à une distribution de Poisson en phase G2

Puisque *S. pombe* passe près de 70% de son cycle cellulaire en phase G2 et 10% dans les phases restantes, nous avons séparé les cellules étudiées en trois différentes phases : G2, M et G1/S (Figure 11). Par défaut, les cellules sont en G2 si celles-ci ne se divisent pas. Si elles le sont, l'absence de septum indiquera qu'elles se trouvent en phase M, alors que la présence du septum indiquera plutôt que celles-ci sont en phase G1/S.

En phase G2, représentée par la Figure 12, la distribution des transcrits naissants ne concorde pas avec une distribution de Poisson.

C'est ainsi dire que même en pleine phase de croissance qu'est la phase de G2, il n'est pas possible de décrire l'expression du gène *mdn1*⁺ comme étant purement stochastique et donc, continu. Par contre, comme le démontre la Figure 12D, lorsque les allèles inactifs ne sont pas pris en considération, la distribution correspond à un processus de Poisson. Cela peut s'expliquer par le fait qu'une fois la transcription initiée, l'élongation se fait de manière stochastique. Certes, le fait que la distribution originale ne corresponde pas à un processus purement stochastique démontre qu'il existe une certaine régulation dans l'expression de *mdn1*⁺ en G2 chez *S. pombe* qui, à ce jour, reste inconnue.

L'expression du gène *mdn1*⁺ ne correspond pas à une distribution de Poisson en phase M, mais le devient en G1/S

Nous avons ensuite voulu suivre l'expression de *mdn1*⁺ en phase M et en G1/S. Nous avons testé les distributions obtenues pour les transcrits naissants de ces deux phases (Figure 13) et avons trouvé que seulement en phase G1/S la transcription de *mdn1*⁺ correspond à une distribution de Poisson.

L'amplitude de la transcription du gène *mdn1*⁺ fluctue selon la phase du cycle cellulaire

Bien que le gène de *mdn1*⁺ ne soit pas un gène caractérisé comme étant régulé par le cycle cellulaire [121, 122, 129], nous avons tout de même observé que son expression

diminue pendant la phase M (Figure 13). À cette phase, le nombre d'allèles inactifs passe de 28% en G2 (Figure 12) à 55% pour retomber à 37% en phase G1/S. Par contre, la transcription de *mdn1*⁺ n'est jamais complètement éteinte (Figure 13).

Afin d'étudier la phase M de manière quantitativement temporelle, nous avons calculé la distance séparant les deux noyaux d'une cellule en division et l'avons divisé par la longueur totale de la cellule. Cela nous donne un coefficient correspondant à une étape de la progression de la phase M du cycle cellulaire avant la septation médiale. Ainsi, nous observons qu'un ralentissement de la transcription du gène *mdn1*⁺ dans *S.pombe* se produit alors que la phase M progresse, sans qu'il y ait un arrêt complet de cette dernière.

Cela pourrait expliquer pourquoi en phase G1/S nous observons une distribution de Poisson. En effet, lorsque la transcription ralentit en phase M, elle reprend en phase G1/S de manière continue et donc aléatoire.

Il a été proposé récemment que le facteur de terminaison TTF2 serait à l'origine d'une répression transcriptionnelle pendant la mitose chez l'humain [130]. Par contre, son homologue chez *S. pombe*, *rrp1*⁺ n'a pas encore été démontré comme ayant cette fonction. Il serait donc intéressant d'étudier ce phénomène afin de mieux le comprendre.

Caractéristiques de l'expression de *mdn1*⁺ dans *S. pombe*

Même si la technique de smFISH ne nous permet pas de suivre une cellule et l'expression de ses gènes en temps réel, elle nous permet tout de même d'étudier l'initiation de la transcription d'un gène.

Dans *S. pombe*, *mdn1*⁺ est un gène ne contenant qu'un seul exon d'une longueur de 14361 paires de bases. Il a été estimé qu'en moyenne chez les levures, un gène est transcrit à une vitesse entre 1kb et 2kb à la minute [131, 132]. Cela veut dire qu'en moyenne, le gène de *mdn1*⁺ prendrait environ 9.6 minutes à être transcrit. Nous avons calculé qu'il y a en moyenne 1.85 transcrits naissants par allèle pendant la phase G2.

Equation 4

$$\text{ARNm/min} = \frac{\ln(2) \times \text{ARNm/cell}}{\text{Demi-vie (min)}}$$

Afin de calculer combien d'ARNm sont produits en moyenne par minutes, nous avons utilisé l'Équation 4. Celle-ci tient en compte le taux de dégradation de l'ARNm, du nombre d'ARNm par cellules retrouvés et du temps de demi-vie de ce dernier. Ainsi, en utilisant la moyenne d'ARNm de *mdn1*⁺ dans *S. pombe* que nous observons en G2 dans nos données, soit de 4.07 transcrits, et du temps de demi-vie de ce dernier par un autre laboratoire, nous trouvons que 4.2 ARNm sont produits par heure.

Pourtant, nous observons en moyenne 1.86 transcrits naissants par allèle. Si le nombre d'ARNm naissants est une mesure temporelle, soit pour *mdn1*⁺ des 9.6 minutes précédents la fixation des cellules, c'est ainsi dire que dans ce laps de temps, 1.86 transcrits ont été produits. Cela signifie qu'un transcrit naissant serait produit à toutes les 5.1 minutes et non à toutes les 14.3 minutes tel que calculé théoriquement.

Chez la levure *S. cerevisiae*, pour le gène de *MDN1*, il avait été calculé qu'environ un transcrit naissant était produit par évènement d'initiation de transcription. Cela est donc devenu un critère pour déterminer qu'un gène est bel et bien transcrit de manière continue, en plus de correspondre à une distribution de Poisson. À la Figure 9A, il est possible de voir que certaines cellules produisent jusqu'à 7 transcrits naissants par allèle, estimant à un nouveau transcrit par 1.4 minutes dans ces cellules. De plus, presque toutes les cellules ont au moins un allèle transcrivant *mdn1*⁺. Ainsi, la transcription du gène *mdn1*⁺ n'est ni purement continue, ni un *burst* robuste. Il est donc possible que sa transcription se produise par *burst* léger surtout en G2, c'est-à-dire que plus d'un transcrit naissant sont produits lors d'un évènement d'initiation, mais il est possible que cela ne soit pas toujours le cas. Ce patron d'expression en *burst* léger n'a pas encore été rapporté dans

la littérature.

Il s'est donc produit, au courant de l'évolution, un mécanisme ou une complexification de la transcription du gène *mdn1*⁺ entre la levure à bourgeonnement et la levure à fission pour que ce dernier ne soit plus exprimé de manière purement aléatoire sauf en phase G1/S. Des études approfondies afin de comprendre les dessous de ce phénomène sont néanmoins nécessaires.

Le patron d'expression du gène *PRP8/spp42*⁺ n'a pas changé au cours de l'évolution entre la levure *S. cerevisiae* et la levure *S. pombe*

Dans la levure *S. cerevisiae*, le gène de *PRP8* a été catégorisé comme un gène s'exprimant par le patron d'expression continu. Nous avons voulu étudier son patron d'expression dans la levure *S. pombe* afin de savoir si ce patron d'expression a changé au courant de l'évolution. Nous avons donc, tout comme pour *mdn1*⁺, quantifié le nombre de transcrits naissants par allèle par smFISH dans une population de cellules *S. pombe* et avons aussi quantifié le nombre de transcrits cytoplasmiques se retrouvant dans cette population.

Les distributions se trouvant à la Figure 15 sont celles obtenues suite à ces quantifications et ont aussi été testé avec le test de χ^2 pour une distribution de Poisson. La distribution des transcrits naissants par allèle a passé le test de χ^2 , contrairement à la distribution des transcrits cytoplasmiques.

Nous avons ensuite voulu savoir si le patron d'expression de *PRP8/spp42*⁺ change selon le stade du cycle cellulaire dans lequel se trouve une cellule. Nous avons donc séparé les cellules ayant été étudié à la Figure 15 selon le stade du cycle cellulaire où chacune d'elle se trouve. Les Figure 17 et 18 représentent les différentes distributions calculées, soit les transcrits naissants et cytoplasmiques à chacun des stades du cycle cellulaire étudiés (G2, M et G1/S). Toutes les distributions de transcrits naissants de la Figure 17 et 18 ont passé le test de χ^2 pour la correspondance avec une distribution de Poisson. Cela signifie que la transcription du gène *PRP8/spp42*⁺ est non seulement en général continue, mais elle l'est tout au long du cycle cellulaire.

Expression de *PRP8/spp42*⁺ pendant la phase M du cycle cellulaire

En observant les distributions de données pour les transcrits naissants en phase M du gène *PRP8/spp42*⁺ (Figure 18A), nous avons remarqué que tout comme pour *mdn1*⁺, il semblerait qu'il y ait un ralentissement dans l'expression de ce dernier. Le nombre d'allèle ne produisant aucun transcrit de *PRP8/spp42*⁺ grimpe drastiquement. Pourtant, ce gène n'est pas connu comme étant régulé par le cycle cellulaire [121, 122]. Nous avons donc calculé, tout comme pour *mdn1*⁺, le coefficient de la distance entre les deux noyaux de chaque cellule et la distance totale de la cellule en division (coefficient d/e). La Figure 19 démontre que contrairement à l'expression de *mdn1*⁺ qui ne fait que diminuer pendant la progression de la phase M, celle de *PRP8/spp42*⁺ s'éteint complètement au début de la phase M et reprend graduellement alors que la phase mitotique prend fin.

Il est donc possible que *PRP8/spp42*⁺ possède une composante régulatrice dans son cycle cellulaire, tout comme *mdn1*⁺. Des études plus approfondies sont nécessaires afin de comprendre s'il s'agit réellement d'une régulation dépendante du cycle cellulaire, ou simplement une répression normale lors de la phase M. De manière intéressante, il est possible d'observer que la répression des deux gènes ne se fait pas en même temps lors de la phase M.

Caractéristiques de l'expression de *PRP8/spp42*⁺ dans *S. pombe*

Dans la levure à fission, le gène *PRP8/spp42*⁺ ne contient qu'un seul exon d'une longueur de 7336 paires de bases. Si l'on estime que dans la levure la vitesse de transcription se situe entre 1kb et 2kb par minute, cela prendrait environ 5 minutes à transcrire le gène au complet. En moyenne, nous observons 1.18 transcrits de *PRP8/spp42*⁺ par allèle dans les cellules en phase G2. Cela signifie qu'en 5 minutes, tout au plus, 1.18 transcrits ont été transcrits. Bien que certaines cellules en transcrivent jusqu'à 4 pendant cette période (2% de la population), la transcription de *PRP8/spp42*⁺ demeure peu fréquente.

En comparant *mdn1*⁺ et *PRP8/spp42*⁺, il est possible de voir qu'un transcrit serait

produit respectivement à toutes les 9.6 et 5 minutes. Par contre, deux options à l'expression se présentent : un transcrit est produit environ toutes les 5-10 minutes ou deux à trois transcrits sont produits dans le même évènement de transcription occasionnellement. La distribution de Poisson correspondrait mieux au premier scénario qui est tout à fait aléatoire et où chaque initiation de transcription est indépendante de la précédente. Le gène *PRP8/spp42⁺* y correspond, et ce, à toutes les phases du cycle cellulaire. Le deuxième scénario pourrait avoir un phénotype très similaire à un processus de Poisson, mais n'y correspondrait mathématiquement pas, ressemblant aux résultats obtenus pour l'expression du gène *mdn1⁺*.

Nous pouvons donc affirmer avec certitude que l'expression du gène *PRP8/spp42⁺* n'a pas changé au cours de l'évolution entre les millions d'années qui séparent les deux levures en question. Par contre, nous suggérons que la transcription de *mdn1⁺* se fait plutôt par *burst* léger dans la levure à fission contrairement à la levure à bourgeonnement.

L'expression des gènes *pol1⁺* et *cdc13⁺* ne correspondent pas à une distribution de Poisson

Dans le but d'étudier le patron d'expression de gènes régulés par le cycle cellulaire, nous avons choisi d'abord le gène *pol1⁺*, encodant pour la sous-unité catalytique de l'ADN polymérase. Ce gène avait été caractérisé comme étant continuellement exprimé dans la levure *S. cerevisiae* [53]. Nous avons aussi étudié un autre gène régulé par le cycle cellulaire, *cdc13⁺* encodant une cycline de type B essentielle dans le passage des cellules de la phase G2 à la phase M du cycle cellulaire. Ce dernier a été étudié plutôt pour sa dégradation lors de la phase M du cycle cellulaire dans la levure *S. cerevisiae*. Il a été montré de fait qu'il n'est pas exprimé en *burst* et qu'il s'accumule plutôt pendant la phase G2 [124, 125].

Nous avons donc utilisé la technique de smFISH afin de quantifier les transcrits naissants et cytoplasmiques de cellules se trouvant d'abord dans toutes les phases du cycle cellulaire (Figure 21). Nous avons évidemment testé par le test statistique de χ^2 si ces

dernières correspondaient à des distributions de Poisson, mais aucune n'y correspondaient.

Puisque ces deux gènes sont connus comme étant régulés par le cycle cellulaire, nous avons étudié leur expression dans toutes les phases cellulaires séparément. La Figure 18 représente les distributions de transcrits naissants par allèle et transcrits cytoplasmiques de *pol1*⁺ et *cdc13*⁺ aux phases G2, M et G1/S. Seule la distribution des transcrits naissants du gène *pol1*⁺ en phase M correspond à une distribution de Poisson. Il est donc possible de conclure que ces gènes ne sont pas transcrits de manière complètement aléatoire.

L'expression des gènes *pol1*⁺ et *cdc13*⁺ est à leur maximum en phase G2

Dans l'étude de Zenklusen et al., le patron d'expression du gène *pol1*⁺ avait été déterminé lorsque l'expression de ce dernier était active. Selon la Figure 22, nous observons que c'est en fait à la phase G2 que l'expression des gènes *pol1*⁺ et *cdc13*⁺ est à son maximum, tant pour les transcrits naissants que pour les transcrits cytoplasmiques. Antérieurement, il a été établi, par des expériences de micropuces sur des souches de *S. pombe* pouvant être arrêtées à une phase du cycle cellulaire spécifique, que l'expression maximale de ces deux gènes serait en phase M [121, 122]. Cette différence peut être principalement expliquée par le fait que nous utilisons des cellules de souche sauvage, où aucune modification génétique n'a été introduite. Ainsi, nous croyons que nos données reflètent mieux l'expression des gènes que nous observons, puisqu'aucun paramètre ne pourrait influencer cette dernière. Il est donc tout à fait plausible que l'expression du gène *cdc13*⁺ atteigne son maximum à la phase G2, puisque ce dernier encode pour une cycline important dans le passage des cellules de la phase G2 à la phase M.

Caractéristiques de l'expression de *pol1*⁺ et *cdc13*⁺ dans *S. pombe*

Le gène *cdc13*⁺ dans *S. pombe* contient deux exons : le premier très court de 232 paires de bases et le deuxième de 2614 paires de bases. Le gène *pol1*⁺, quant à lui, possède deux exons respectivement de 428 et 4760 paires de bases. Nous estimons donc, qu'à une vitesse moyenne de 1.5kb/min, la transcription de ces derniers devrait se faire respectivement en 2.13 minutes et 3.5 minutes.

Pour le gène *pol1*⁺ et *cdc13*⁺ respectivement, en G2 en moyenne 0.77 et 1.34 transcrits sont trouvés par allèle, baissant à 0.55 et à 0.73 transcrits par allèle en phase M et à 0.48 et 0.8 transcrits par allèle en G1/S.

Ainsi, au plus fort de leur expression, un transcrit de *pol1*⁺ sera produit à toutes les 4.6 minutes et un transcrit de *cdc13*⁺ produit à toutes les 1.6 minutes environ. Nous ne pouvons donc pas parler ici d'un *burst* robuste comme patron d'expression, ni de patron continué puisque nos distributions ne sont pas aléatoires comme un processus de Poisson. Il est donc possible que *pol1*⁺ et *cdc13*⁺ suivent le même type d'expression que celui du gène *mdn1*⁺, soit d'un *burst* léger où quelques transcrits (e.g. entre 1 et 3) peuvent être produits dans le même évènement de transcription.

Comparaison de l'expression des gènes *mdn1*⁺, *PRP8/spp42*⁺, *pol1*⁺ et *cdc13*⁺ en G2 dans la levure à fission

À la lumière de nos résultats, nous avons voulu comparer nos résultats de l'expression de *mdn1*⁺ obtenus dans *S. pombe* avec ceux antérieurement obtenus dans *S. cerevisiae* [53]. Nous observons qu'une plus grande proportion de cellules ne contient aucun transcrit de *mdn1*⁺ dans *S. pombe* (Figure 26 rouge) par rapport à *S. cerevisiae* (Figure 26 bleu). Quoique ces deux distributions ne se superposent pas parfaitement, elles se suivent quelque peu et suggèrent que le changement évolutif dans l'expression de *mdn1*⁺ ne s'est pas fait de manière drastique entre la levure à bourgeonnement et la levure à fission.

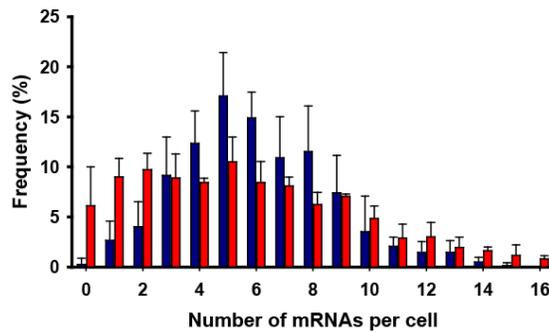


Figure 26. Distributions des fréquences d'ARNm par cellules dans *S. cerevisiae* (bleu) provenant de [53] et dans *S. pombe* (rouge) obtenue dans le présent projet.

Nous avons enfin voulu comparer l'expression de chacun des gènes d'intérêt, et ce, en phase G2 du cycle cellulaire, puisqu'il s'agit de la phase où ces derniers sont le plus transcrits. Le patron d'expression continuuel a été défini comme étant aléatoire suivant une distribution de Poisson. Il a été proposé que pour ce patron d'expression, le promoteur d'un gène soit constamment « ouvert » et où ce sont les initiations de transcription qui soient indépendantes les unes des autres (Figure 27A). Dans le cas du patron d'expression en *burst*, le promoteur passe de l'état actif à inactif, et lors de l'activation de ce dernier, de multiples transcrits sont produits à la fois. Ce dernier mode correspond à une distribution de fréquence bimodale (Figure 27B). Dans le présent projet, nous proposons un mode d'expression intermédiaire entre ces deux extrêmes (Figure 27C). Nous avons pu observer ce dernier dans certains des gènes que nous avons étudiés. Le seul gène étudié correspondant parfaitement à une expression en continu est *PRP8/spp42*⁺. Toutes les distributions des transcrits naissants, peu importe la phase du cycle cellulaire, correspondent à une distribution de Poisson. Par contre, les gènes *mdn1*⁺, *pol1*⁺ et *cdc13*⁺ ne correspondaient pas à une distribution de Poisson en phase G2.

Il est à noter que la variabilité dans nos distributions cytoplasmiques n'a pas été au centre de notre modèle puisque nous nous intéressons précisément à l'initiation de la transcription. Ainsi, la durée de vie des ARNm cytoplasmiques ne pourrait être prise en compte dans ce dernier. Nous croyons qu'il est plus probable que cette variabilité dans les

distributions cytoplasmiques soit d'origine biologique, puisque l'intensité de chacun des transcrits détectés était similaire. Cela nous permet d'en déduire que la variabilité n'était pas de nature expérimentale.

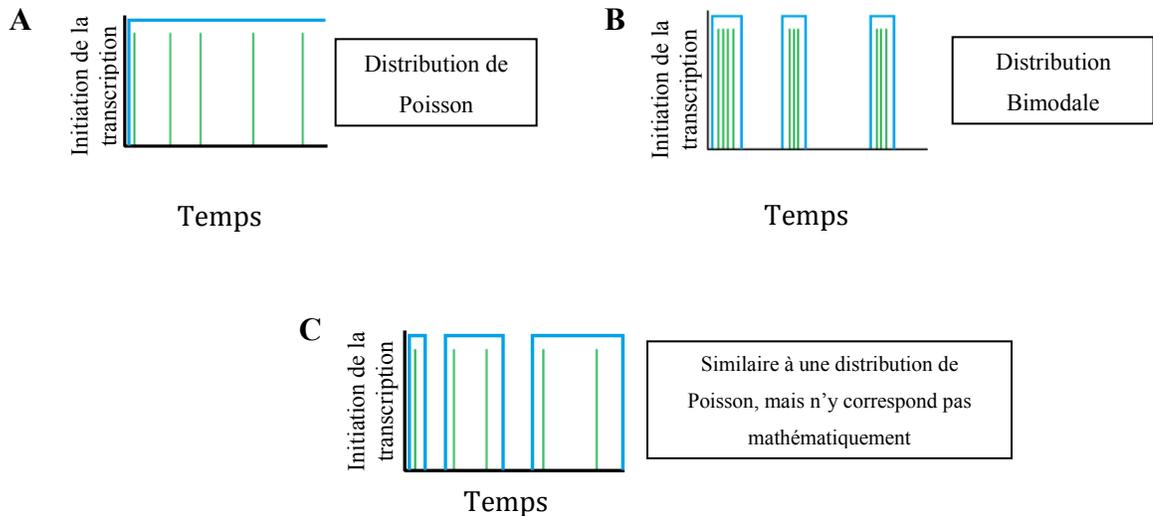


Figure 27. Modèles d'initiation de la transcription avec l'activation du promoteur (ligne bleue) et l'initiation de la transcription est représentée par les barres verticales verte. **A.** Patron d'expression continu où le promoteur est actif de manière continue et où la transcription est initiée de manière aléatoire suivant une distribution de Poisson. **B.** Patron d'expression en *burst*, où le temps d'inactivité du promoteur est grand et où le nombre de transcrits naissants par période d'activité est élevé correspondant à une distribution bimodale. **C.** Modèle de *burst* léger où le promoteur possède des périodes d'inactivités et où, par initiation, plus d'un transcrit est produit. Ce modèle ne mène ni à une distribution de Poisson, ni à une distribution bimodale.

Pour ces gènes, nous avons comparé leur taux de transcription théorique et observé dans nos expériences (Tableau 4). Par exemple, nous avons calculé que le taux de transcription théorique (calculé à partir de l'Équation 4) est trois fois moins élevé que celui observé aux sites de transcription de chaque allèle transcrivant *mdn1*⁺ en G2 dans *S. pombe*. Pour le gène *PRP8/spp42*⁺, le taux de transcription théorique est près de deux fois plus petit que le taux de transcription observé. Pour le gène *cdc13*⁺, il s'agit d'environ le même rapport. Par contre, le taux de transcription du gène *poll*⁺ observé dans les allèles n'est pas loin du taux de transcription théorique. Ce phénomène peut être expliqué par le

fait que la vitesse de transcription de ces gènes est supérieure à celle utilisée dans nos calculs. Il est aussi possible que le temps de demi-vie utilisé dans l'Équation 4 (calcul du taux de transcription théorique) pour les gènes *mdn1*⁺, *PRP8/spp42*⁺ et *cdc13*⁺ soient inférieurs à la réalité. Il faudrait donc calculer le temps de demi-vie des ARNm, par exemple par imagerie en temps réel et en utilisant des inhibiteurs de la transcription.

Quoi que les distributions de fréquence des transcrits naissants en G2 des gènes *mdn1*⁺ et *pol1*⁺ ne correspondent pas à une distribution de Poisson, leur taux de transcription est très similaire à celui de *PRP8/spp42*⁺, qui est exprimé continuellement. Pour ce qui est du gène *cdc13*⁺, un taux de transcription de 38 transcrits par heure ne peut directement être expliqué par le patron d'expression en *burst*.

Nous avons donc calculé pour chacun des gènes le facteur Fano, c'est-à-dire la mesure de dispersion par rapport à une distribution de Poisson ayant la même moyenne que la distribution testée. Une distribution de Poisson a un facteur Fano de 1. Nous pouvons observer d'abord que le facteur Fano de *cdc13*⁺ n'est pas très loin de 1, ce qui nous permet de suggérer que, tout comme *mdn1*⁺ et *pol1*⁺, *cdc13*⁺ est exprimé selon un patron intermédiaire (Figure 27C).

Une observation intéressante est que le seul facteur Fano supérieur à 1 est celui du gène *mdn1*⁺. Dans une récente étude, un groupe a étudié les facteurs Fano de plusieurs gènes dans la levure *S. cerevisiae*, et ont proposé un modèle selon lequel la vitesse d'amarrage de l'ARN Polymérase II n'est pas négligeable par rapport à la vitesse d'élongation (modèle à deux étapes) [133]. En fait, dans l'étude du groupe de Robert H. Singer, on avait supposé que la vitesse d'initiation et la vitesse d'élongation était la même (modèle à une étape). Ainsi, Choubey et al. ont repris des données d'études antérieures et ont tenté de trouver une corrélation entre le facteur Fano et les modèles à une ou deux étapes. Ils ont trouvé que dans *S. cerevisiae*, lorsque le facteur Fano est très près de 1, il s'agit du modèle à une étape et s'il y est inférieur, il s'agirait plutôt du modèle à deux étapes. Enfin, si le facteur Fano est supérieur à 1, comme dans le cas du gène PDR5 (Fano ~2), il s'agirait du modèle d'expression en *burst*.

Suivant donc ces observations, le gène *PRP8/spp42*⁺ aurait un modèle en une étape, les gènes *pol1*⁺ et *cdc13*⁺ en deux étapes, et *mdn1*⁺ serait plutôt exprimé en *burst*. Il faut par contre noter que les modèles à une ou deux étapes sont continus.

Gène	Longueur (pb)	Temps de transcription (min)	Taux de transcription théorique (/min)	Taux de transcription théorique (/heure)	Nombre de naissants théoriques	Moyenne des naissants observée	Moyenne de naissants par heure observée	Facteur Fano en G2
<i>mdn1</i> ⁺	14360	9.57	0.07	4.2	0.67	1.86	11.66	1.4986
<i>prp8</i>	7340	4.89	0.14	8.4	0.69	1.18	14.47	0.9994
<i>cdc13</i> ⁺	3200	2.13	0.27	16.2	0.58	1.34	37.69	0.938
<i>pol1</i> ⁺	5280	3.52	0.19	11.4	0.67	0.77	13.13	0.8147

Tableau 4. Tableau comparatif des différents paramètres calculés théoriques et observés pour chacun des gènes étudiés en G2 dans la levure *S. pombe* tels que la longueur, le temps de transcription des gènes, le taux théorique de transcription des gènes par minute et par heure et les moyennes des transcrits naissants observée et le facteur Fano de chaque gène en G2.

À ce jour, il ne nous est pas possible de distinguer le patron d'expression en continu d'un patron d'expression en *burst* léger, sauf en déterminant si la distribution des fréquences de transcrits naissants correspond à une distribution de Poisson. Puisque nous n'observons pas de distributions bimodales, signe que les gènes étudiés sont exprimés par *burst*, nous ne pouvons conclure que *mdn1*⁺, *pol1*⁺ et *cdc13*⁺ soient exprimés en *burst* robuste.

Il serait possible d'étudier la cinétique de l'initiation de la transcription par imagerie en temps réel, soit par le système PP7/MS2 [108, 134]. Nous avons de fait étiqueté le gène *mdn1*⁺ dans la levure *S. pombe* avec le système PP7, mais n'avons pas obtenu de résultats concluants. En effet, nous avons remarqué plusieurs agrégats dans les cellules observées. Aussi, les cellules semblaient ne pas croître comme les cellules de souche sauvage. Ainsi, ce système n'est pas tout à fait au point dans la levure *S. pombe* et nécessite plusieurs étapes

d'optimisation. Par contre, l'avantage d'un système pareil réside dans le fait qu'il est même possible d'observer simultanément la production des ARNm et des protéines encodés. Nous travaillons présentement au laboratoire afin de peaufiner cette technique qui permettra enfin de voir si les variations transcriptionnelles et traductionnelles sont intimement reliées.

Résultats de smFISH dans des lignées cellulaires humaines

Suite à nos observations dans la levure *Schizosaccharomyces pombe*, nous avons voulu étudier les mêmes gènes dans des lignées cellulaires humaines. Nous avons débuté par deux lignées cellulaires cancéreuses, soit la lignée HEK293T provenant de reins embryonnaires et U2OS étant des cellules d'ostéosarcome (Figure 24 ABC (HEK293T), DEF (U2OS)). Nous avons aussi voulu étudier l'expression des gènes dans une lignée cellulaire primaire, IMR 90, provenant de fibroblastes humains. Cette dernière lignée cellulaire, quoique plus compliquée à cultiver, donne des résultats beaucoup plus précis sur l'expression des gènes au niveau physiologique, puisque non cancéreuse.

Nous avons d'abord commencé par cibler les exons de l'extrémité 5' du gène *mdn1*⁺ par smFISH dans des cellules HEK293T. Nous avons d'abord remarqué plusieurs points très intenses dans les cellules ainsi que plusieurs ARNm nucléaires. Puisque les cellules HEK293T sont triploïdes, et qu'il est possible que certains des allèles soient inactifs, il était difficile d'identifier les sites de transcription du gène afin de le quantifier (Figure 24 A, D, G). Nous avons donc ciblé dans les cellules, utilisant un fluorophore différent de celui des sondes en 5', l'extrémité des exons en 3' du gène *mdn1*⁺. Ainsi, nous pouvons comparer les signaux des deux extrémités et/ou établir que si des signaux nucléaires 5' ne colocalisent pas avec un signal 3', ils seront considérés comme potentiels sites de transcription.

Nous observons que, dans les cellules HEK293T, les sites de transcription de MDN1 sont plus intenses que dans les cellules U2OS ou IMR90 où les sites de transcription discrets pour le gène MDN1. L'expression génique en général est plus forte dans les cellules HEK293T dû à leur métabolisme typiquement plus actif. Puisque nous observons souvent trois intenses sites de transcription pour le gène MDN1 dans les cellules HEK293T, cela

suggère que la transcription est active dans presque tous les allèles présents, ne correspondant pas au patron d'expression en *burst*. Les sites de transcription plutôt faibles observés dans les cellules U2OS et IMR90 rendent difficile la spéculation quant au patron d'expression de MDN1 dans ces dernières. Il est donc possible que même dans les cellules humaines, le gène MDN1 ne soit pas exprimé en *burst* robuste, mais des analyses quantitatives sont nécessaires afin de conclure cela. Comme décrit précédemment, dans la section Résultats, nous avons mis au point des outils bio-informatiques qui nous permettront de déterminer de manière efficace le patron d'expression des gènes dans des lignées cellulaires humaines (Figure 25).

Nous avons fait, lors de l'acquisition d'images des transcrits de MDN1 par smFISH de ces lignées cellulaires, une observation intéressante : la plupart des sites de transcription contiennent un signal 3'. Cela nous a interpellés puisque les sondes à l'extrémité 5' s'étendent sur les 60763 premiers nucléotides de l'ARNm, alors que ceux de l'extrémité 3' s'étendent sur les 16941 derniers nucléotides. Au total, l'ARNm du gène MDN1 fait 177224 nucléotides. Si la vitesse de transcription par l'ARN Polymérase II est constant et d'environ 30 nucléotides/seconde, cela prendrait 5907.5 secondes, soit 98.46 minutes à transcrire MDN1. La transcription de l'extrémité 5' prendrait 33.8 minutes à transcrire, et 9.4 minutes pour l'extrémité 3'. Nous avons calculé que si la vitesse est constante, il y a environ 4% de probabilité qu'un signal 3' soit visible là où un ou deux signaux 5' le sont. D'après nos observations, la plupart des sites de transcription possèdent un signal 3', soit plus de 4% des cas. Il est donc possible que les ARNm restent un long moment avant d'être relâché des sites de transcription, soit pour une étape de contrôle de qualité ou simplement que la terminaison de la transcription se ferait lentement.

E. Conclusion et perspectives

Afin de comprendre si les patrons d'expression de gène sont gène ou organisme spécifiques, nous avons étudié par la technique de smFISH quatre gènes évolutivement conservés. Ils ont tous été précédemment caractérisés comme étant exprimés de manière continue dans la levure *S. cerevisiae*, nous les avons donc étudiés d'abord dans la levure *S. pombe*.

Nos résultats ont pu démontrer que seul le gène *PRP8/spp42*⁺ respecte les conditions d'expression par le patron continu dans la levure *S. pombe*. Par contre, nous n'avons pas pu catégoriser les gènes restants comme *bursting* puisqu'aucune des distributions de fréquence de transcrits n'arborait une distribution bimodale. De fait, *mdn1*⁺ est exprimé continuellement en phase G1/S et *poll*⁺ l'est aussi pendant la phase M. En les comparant les uns aux autres, nous avons pu observer que leur expression est similaire à celle de *PRP8/spp42*⁺, même si elle ne correspond pas mathématiquement à une distribution de Poisson. Ainsi, nous pouvons conclure que l'expression du gène *PRP8/spp42*⁺ n'a pas changé au cours de l'évolution entre la levure *S. cerevisiae* et la levure *S. pombe*. Par contre, il est possible que les autres gènes soient exprimés en légers *burst* dans la levure *S. pombe*, ce qui constituerait un changement évolutif dans le mode d'expression de ces gènes. Ainsi, les patrons d'expression seraient **gène et organisme spécifiques**.

Enfin, nous n'avons pu qu'obtenir des résultats préliminaires dans trois lignées cellulaires humaines, dont deux sont cancéreuses et l'une est primaire. Nos observations suggèrent que le patron d'expression de *mdn1*⁺ ne serait pas en *burst*, mais des analyses quantitatives seront nécessaires afin de déterminer le patron d'expression de *mdn1*⁺ dans les cellules humaines.

Initialement, nous avons tenté de déterminer l'expression des gènes présentés dans plusieurs autres organismes à des stades de l'évolution et de complexité différents, tels que dans le vers *C. elegans* et dans la mouche *D. melanogaster*. Cette étude représente un

premier pas vers la compréhension de l'évolution des patrons d'expression de gènes et nous avons pu développer des outils essentiels afin de poursuivre cette dernière.

Bibliographie

1. Shen, X., et al., *A chromatin remodelling complex involved in transcription and DNA processing*. Nature, 2000. **406**(6795): p. 541-4.
2. Wilson, C.J., et al., *RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling*. Cell, 1996. **84**(2): p. 235-44.
3. Martens, J.A. and F. Winston, *Recent advances in understanding chromatin remodeling by Swi/Snf complexes*. Curr Opin Genet Dev, 2003. **13**(2): p. 136-42.
4. Buratowski, S., et al., *Five intermediate complexes in transcription initiation by RNA polymerase II*. Cell, 1989. **56**(4): p. 549-61.
5. Kim, Y.J., et al., *A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II*. Cell, 1994. **77**(4): p. 599-608.
6. Koleske, A.J. and R.A. Young, *An RNA polymerase II holoenzyme responsive to activators*. Nature, 1994. **368**(6470): p. 466-469.
7. Ossipow, V., et al., *A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation*. Cell, 1995. **83**(1): p. 137-146.
8. Cairns, B.R., et al., *RSC, an essential, abundant chromatin-remodeling complex*. Cell, 1996. **87**(7): p. 1249-1260.
9. Chao, D.M. and R.A. Young, *Activation without a vital ingredient*. Nature, 1996. **383**(6596): p. 119-120.
10. Maldonado, E., et al., *A human RNA polymerase II complex associated with SRB and DNA-repair proteins*. Nature, 1996. **381**(6577): p. 86-89.
11. Nakajima, T., et al., *RNA helicase A mediates association of CBP with RNA polymerase II*. Cell, 1997. **90**(6): p. 1107-1112.
12. Wu, S.Y., et al., *Isolation of mouse TFIID and functional characterization of TBP and TFIID in mediating estrogen receptor and chromatin transcription*. J Biol Chem, 1999. **274**(33): p. 23480-90.
13. Wu, C.H., et al., *NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila*. Genes Dev, 2003. **17**(11): p. 1402-14.
14. Liu, Y., et al., *Yeast nuclear extract contains two major forms of RNA polymerase II mediator complexes*. J Biol Chem, 2001. **276**(10): p. 7169-75.
15. He, Y., et al., *Structural visualization of key steps in human transcription initiation*. Nature, 2013. **495**(7442): p. 481-6.
16. Thomas, M.C. and C.M. Chiang, *The general transcription machinery and general cofactors*. Crit Rev Biochem Mol Biol, 2006. **41**(3): p. 105-78.
17. Harmston, N. and B. Lenhard, *Chromatin and epigenetic features of long-range gene regulation*. Nucleic acids research, 2013. **41**(15): p. 7185-7199.
18. Guarente, L., *UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals*. Cell, 1988. **52**(3): p. 303-5.
19. Nobrega, M., et al., *Scanning human gene deserts for long-range enhancers*. Science, 2003. **302**(5644): p. 413-413.
20. Levine, M. and R. Tjian, *Transcription regulation and animal diversity*. Nature, 2003. **424**(6945): p. 147-51.
21. Kim, Y.J., et al., *A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II*. Cell, 1994. **77**(4): p. 599-608.
22. Goldman, S.R., R.H. Ebright, and B.E. Nickels, *Direct Detection of Abortive RNA Transcripts in Vivo*. Science, 2009. **324**(5929): p. 927-928.
23. Pokholok, D.K., N.M. Hannett, and R.A. Young, *Exchange of RNA Polymerase II Initiation and Elongation Factors during Gene Expression In Vivo*. Molecular Cell, 2002. **9**(4): p. 799-809.
24. Wade, J.T. and K. Struhl, *The transition from transcriptional initiation to elongation*. Current Opinion in Genetics & Development, 2008. **18**(2): p. 130-136.
25. Thomas, M.J., A.A. Platas, and D.K. Hawley, *Transcriptional fidelity and proofreading by RNA polymerase II*. Cell, 1998. **93**(4): p. 627-637.
26. Wang, D., et al., *Structural Basis of Transcription: Backtracked RNA Polymerase II at 3.4 Angstrom Resolution*. Science, 2009. **324**(5931): p. 1203-1206.
27. Saunders, A., L. Core, and J. Lis, *Breaking barriers to transcription elongation*. Nature reviews. Molecular cell biology, 2006. **7**(8): p. 557-567.

28. Brès, V., S.M. Yoh, and K.A. Jones, *The multi-tasking P-TEFb complex*. Current Opinion in Cell Biology, 2008. **20**(3): p. 334-340.
29. Mayer, A., et al., *CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II*. Science, 2012. **336**(6089): p. 1723-1725.
30. Whitelaw, E. and N. Proudfoot, *Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene*. EMBO journal, 1986. **5**(11): p. 2915-2922.
31. Birse, C.E., et al., *Coupling termination of transcription to messenger RNA maturation in yeast*. Science, 1998. **280**(5361): p. 298-301.
32. Yonaha, M. and N.J. Proudfoot, *Transcriptional termination and coupled polyadenylation in vitro*. EMBO journal, 2000. **19**(14): p. 3770-3777.
33. Proudfoot, N., *New perspectives on connecting messenger RNA 3' end formation to transcription*. Current opinion in cell biology, 2004. **16**(3): p. 272-278.
34. Hirose, Y. and J.L. Manley, *RNA polymerase II and the integration of nuclear events*. Genes & development, 2000. **14**(12): p. 1415-1429.
35. Connelly, S. and J.L. Manley, *A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II*. Genes & development, 1988. **2**(4): p. 440-452.
36. Logan, J., et al., *A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene*. Proceedings of the National Academy of Sciences of the United States of America, 1987. **84**(23): p. 8306-8310.
37. Richard, P. and J.L. Manley, *Transcription termination by nuclear RNA polymerases*. Genes Dev, 2009. **23**(11): p. 1247-69.
38. Rosonina, E., S. Kaneko, and J.L. Manley, *Terminating the transcript: breaking up is hard to do*. Genes Dev, 2006. **20**(9): p. 1050-6.
39. Hara, R., et al., *Human transcription release factor 2 dissociates RNA polymerases I and II stalled at a cyclobutane thymine dimer*. J Biol Chem, 1999. **274**(35): p. 24779-86.
40. Schubert, C., *Single-cell analysis: The deepest differences*. Nature, 2011. **480**(7375): p. 133-7.
41. Ozbudak, E.M., et al., *Regulation of noise in the expression of a single gene*. Nat Genet, 2002. **31**(1): p. 69-73.
42. Elowitz, M.B., et al., *Stochastic gene expression in a single cell*. Science, 2002. **297**(5584): p. 1183-6.
43. Blake, W.J., et al., *Noise in eukaryotic gene expression*. Nature, 2003. **422**(6932): p. 633-7.
44. Raser, J.M. and E.K. O'Shea, *Control of stochasticity in eukaryotic gene expression*. Science, 2004. **304**(5678): p. 1811-4.
45. Kaern, M., et al., *Stochasticity in gene expression: from theories to phenotypes*. Nat Rev Genet, 2005. **6**(6): p. 451-64.
46. Shahrezaei, V., J.F. Ollivier, and P.S. Swain, *Colored extrinsic fluctuations and stochastic gene expression*. Mol Syst Biol, 2008. **4**: p. 196.
47. Rosenfeld, N., et al., *Gene regulation at the single-cell level*. Science, 2005. **307**(5717): p. 1962-5.
48. Volfson, D., et al., *Origins of extrinsic variability in eukaryotic gene expression*. Nature, 2006. **439**(7078): p. 861-4.
49. Golding, I., et al., *Real-time kinetics of gene activity in individual bacteria*. Cell, 2005. **123**(6): p. 1025-1036.
50. Cai, L., N. Friedman, and X.S. Xie, *Stochastic protein expression in individual cells at the single molecule level*. Nature, 2006. **440**(7082): p. 358-362.
51. Taniguchi, Y., et al., *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells*. Science, 2010. **329**(5991): p. 533-8.
52. Komorowski, M., J. Miekisz, and A. Kierzek, *Translational repression contributes greater noise to gene expression than transcriptional repression*. Biophysical journal, 2009. **96**(2): p. 372-384.
53. Zenklusen, D., D. Larson, and R. Singer, *Single-RNA counting reveals alternative modes of gene expression in yeast*. Nature structural & molecular biology, 2008. **15**(12): p. 1263-1271.

54. Blake, W.J., et al., *Phenotypic consequences of promoter-mediated transcriptional noise*. Mol Cell, 2006. **24**(6): p. 853-65.
55. Munsky, B., G. Neuert, and A. van Oudenaarden, *Using gene expression noise to understand gene regulation*. Science, 2012. **336**(6078): p. 183-7.
56. Suter, D., et al., *Mammalian genes are transcribed with widely different bursting kinetics*. Science, 2011. **332**(6028): p. 472-474.
57. Muramoto, T., et al., *Live imaging of nascent RNA dynamics reveals distinct types of transcriptional pulse regulation*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(19): p. 7350-7355.
58. Reinke, H. and D. Gatfield, *Genome-wide oscillation of transcription in yeast*. Trends in biochemical sciences, 2006. **31**(4): p. 189-191.
59. Raj, A., et al., *Stochastic mRNA synthesis in mammalian cells*. Plos Biology, 2006. **4**(10): p. 1707-1719.
60. Shahrezaei, V. and P.S. Swain, *The stochastic nature of biochemical networks*. Curr Opin Biotechnol, 2008. **19**(4): p. 369-74.
61. Suter, D., et al., *Origins and consequences of transcriptional discontinuity*. Current opinion in cell biology, 2011. **23**(6): p. 657-662.
62. Butte, A.J., V.J. Dzau, and S.B. Glueck, *Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues"*. Physiological genomics, 2001. **7**(2): p. 95-96.
63. Eisenberg, E. and E. Levanon, *Human housekeeping genes are compact*. Trends in genetics, 2003. **19**(7): p. 362-365.
64. Skupsky, R., et al., *HIV promoter integration site primarily modulates transcriptional burst size rather than frequency*. PLoS Comput Biol, 2010. **6**(9).
65. Chubb, J.R., et al., *Transcriptional pulsing of a developmental gene*. Current Biology, 2006. **16**(10): p. 1018-1025.
66. Boettiger, A.N. and M. Levine, *Synchronous and stochastic patterns of gene activation in the Drosophila embryo*. Science, 2009. **325**(5939): p. 471-3.
67. Becskei, A., B.B. Kaufmann, and A. van Oudenaarden, *Contributions of low molecule number and chromosomal positioning to stochastic gene expression*. Nat Genet, 2005. **37**(9): p. 937-44.
68. Cook, P.R., *A model for all genomes: the role of transcription factories*. J Mol Biol, 2010. **395**(1): p. 1-10.
69. Consortium, E.P., *The ENCODE (ENCyclopedia Of DNA Elements) Project*. Science, 2004. **306**(5696): p. 636-40.
70. Aida, M., et al., *Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene*. Mol Cell Biol, 2006. **26**(16): p. 6094-104.
71. Sims, R.J., 3rd, R. Belotserkovskaya, and D. Reinberg, *Elongation by RNA polymerase II: the short and long of it*. Genes Dev, 2004. **18**(20): p. 2437-68.
72. Sims, R.J., 3rd, S.S. Mandal, and D. Reinberg, *Recent highlights of RNA-polymerase-II-mediated transcription*. Curr Opin Cell Biol, 2004. **16**(3): p. 263-71.
73. Giardina, C., M. Prez-Riba, and J.T. Lis, *Promoter melting and TFIID complexes on Drosophila genes in vivo*. Genes & development, 1992. **6**(11): p. 2190-2200.
74. Yunger, S., et al., *Single-allele analysis of transcription kinetics in living mammalian cells*. Nature Methods, 2010. **7**(8): p. 631-633.
75. Newman, J.R. and J.S. Weissman, *Systems biology: many things from one*. Nature, 2006. **444**(7119): p. 561-2.
76. Murphy, K., et al., *Tuning and controlling gene expression noise in synthetic gene networks*. Nucleic acids research, 2010. **38**(8): p. 2712-2726.
77. Sanchez, A., S. Choubey, and J. Kondev, *Regulation of noise in gene expression*. Annu Rev Biophys, 2013. **42**: p. 469-91.
78. O'Sullivan, J.M., et al., *Gene loops juxtapose promoters and terminators in yeast*. Nat Genet, 2004. **36**(9): p. 1014-8.

79. Anderson, L.M. and H. Yang, *DNA looping can enhance lysogenic CI transcription in phage lambda*. Proc Natl Acad Sci U S A, 2008. **105**(15): p. 5827-32.
80. Yun, K., et al., *Lymphoid enhancer binding factor 1 regulates transcription through gene looping*. J Immunol, 2009. **183**(8): p. 5129-37.
81. O'Reilly, D. and D.R. Greaves, *Cell-type-specific expression of the human CD68 gene is associated with changes in Pol II phosphorylation and short-range intrachromosomal gene looping*. Genomics, 2007. **90**(3): p. 407-15.
82. Perkins, K.J., et al., *Transcription-dependent gene looping of the HIV-1 provirus is dictated by recognition of pre-mRNA processing signals*. Mol Cell, 2008. **29**(1): p. 56-68.
83. Hebenstreit, D., *Are gene loops the cause of transcriptional noise?* Trends Genet, 2013.
84. Roy, S.W. and W. Gilbert, *The evolution of spliceosomal introns: patterns, puzzles and progress*. Nat Rev Genet, 2006. **7**(3): p. 211-21.
85. Rodriguez-Trelles, F., R. Tarrío, and F.J. Ayala, *Origins and evolution of spliceosomal introns*. Annu Rev Genet, 2006. **40**: p. 47-76.
86. Logsdon, J.M., Jr., A. Stoltzfus, and W.F. Doolittle, *Molecular evolution: recent cases of spliceosomal intron gain?* Curr Biol, 1998. **8**(16): p. R560-3.
87. Gibbs, R.A., et al., *Genome sequence of the Brown Norway rat yields insights into mammalian evolution*. Nature, 2004. **428**(6982): p. 493-521.
88. Herrero, E., M.a.A. de la Torre, and E. Valentín, *Comparative genomics of yeast species: new insights into their biology*. International microbiology, 2003. **6**(3): p. 183-190.
89. Forsburg, S.L., *The yeasts Saccharomyces cerevisiae and Schizosaccharomyces pombe: models for cell biology research*. Gravit Space Biol Bull, 2005. **18**(2): p. 3-9.
90. Fantes, P. and P. Nurse, *Control of cell size at division in fission yeast by a growth-modulated size control over nuclear division*. Exp Cell Res, 1977. **107**(2): p. 377-86.
91. Fantes, P.A., *Control of cell size and cycle time in Schizosaccharomyces pombe*. J Cell Sci, 1977. **24**: p. 51-67.
92. Sommer, R. and A. Streit, *Comparative genetics and genomics of nematodes: genome structure, development, and lifestyle*. Annual review of genetics, 2011. **45**: p. 1-20.
93. Adams, M.D., et al., *The genome sequence of Drosophila melanogaster*. Science, 2000. **287**(5461): p. 2185-95.
94. Balazsi, G., A. van Oudenaarden, and J.J. Collins, *Cellular decision making and biological noise: from microbes to mammals*. Cell, 2011. **144**(6): p. 910-25.
95. Raj, A., et al., *Imaging individual mRNA molecules using multiple singly labeled probes*. Nat Methods, 2008. **5**(10): p. 877-9.
96. Walters, M.C., et al., *Enhancers increase the probability but not the level of gene expression*. Proc Natl Acad Sci U S A, 1995. **92**(15): p. 7125-9.
97. Wernet, M., et al., *Stochastic spineless expression creates the retinal mosaic for colour vision*. Nature, 2006. **440**(7081): p. 174-180.
98. Wardle, F.C. and J.C. Smith, *Refinement of gene expression patterns in the early Xenopus embryo*. Development, 2004. **131**(19): p. 4687-96.
99. van Roon, M.A., et al., *The initiation of hepatocyte-specific gene expression within embryonic hepatocytes is a stochastic event*. Dev Biol, 1989. **136**(2): p. 508-16.
100. Kurakin, A., *Self-organization vs Watchmaker: stochastic gene expression and cell differentiation*. Dev Genes Evol, 2005. **215**(1): p. 46-52.
101. Simpson, P., *Notch signalling in development: on equivalence groups and asymmetric developmental potential*. Curr Opin Genet Dev, 1997. **7**(4): p. 537-42.
102. Hume, D.A., *Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression*. Blood, 2000. **96**(7): p. 2323-8.
103. Bauman, J.G., et al., *A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA*. Exp Cell Res, 1980. **128**(2): p. 485-90.
104. Singer, R.H. and D.C. Ward, *Actin gene expression visualized in chicken muscle tissue culture by*

- using *in situ* hybridization with a biotinated nucleotide analog. Proc Natl Acad Sci U S A, 1982. **79**(23): p. 7331-5.
105. Femino, A.M., et al., *Visualization of single RNA transcripts in situ*. Science, 1998. **280**(5363): p. 585-90.
 106. Batish, M., A. Raj, and S. Tyagi, *Single molecule imaging of RNA in situ*. Methods Mol Biol, 2011. **714**: p. 3-13.
 107. Rahman, S. and D. Zenklusen, *Single-molecule resolution fluorescent in situ hybridization (smFISH) in the yeast S. cerevisiae*. Methods Mol Biol, 2013. **1042**: p. 33-46.
 108. Hocine, S., et al., *Single-molecule analysis of gene expression using two-color RNA labeling in live yeast*. Nature Methods, 2013. **10**(2): p. 119-121.
 109. Zenklusen, D. and R.H. Singer, *Analyzing mRNA expression using single mRNA resolution fluorescent in situ hybridization*. Methods Enzymol, 2010. **470**: p. 641-59.
 110. Carpenter, A.E., et al., *CellProfiler: image analysis software for identifying and quantifying cell phenotypes*. Genome Biol, 2006. **7**(10): p. R100.
 111. Lamprecht, M.R., D.M. Sabatini, and A.E. Carpenter, *CellProfiler: free, versatile software for automated biological image analysis*. Biotechniques, 2007. **42**(1): p. 71-5.
 112. Wood, V., et al., *PomBase: a comprehensive online resource for fission yeast*. Nucleic Acids Res, 2012. **40**(Database issue): p. D695-9.
 113. Kersey, P.J., et al., *Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species*. Nucleic Acids Res, 2012. **40**(Database issue): p. D91-7.
 114. Kersey, P.J., et al., *Ensembl Genomes 2013: scaling up access to genome-wide data*. Nucleic Acids Res, 2013.
 115. Treck, T., et al., *Single-mRNA counting using fluorescent in situ hybridization in budding yeast*. Nat Protoc, 2012. **7**(2): p. 408-19.
 116. Garbarino, J.E. and I.R. Gibbons, *Expression and genomic analysis of midasin, a novel and highly conserved AAA protein distantly related to dynein*. BMC Genomics, 2002. **3**: p. 18.
 117. Miyata, M. and H. Miyata, *Relationship between extracellular enzymes and cell growth during the cell cycle of the fission yeast Schizosaccharomyces pombe: acid phosphatase*. J Bacteriol, 1978. **136**(2): p. 558-64.
 118. Raj, A. and A. van Oudenaarden, *Single-molecule approaches to stochastic gene expression*. Annual review of biophysics, 2009. **38**: p. 255-270.
 119. Wills, Q., et al., *Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments*. Nature biotechnology, 2013. **31**(8): p. 748-752.
 120. Grainger, R.J. and J.D. Beggs, *Prp8 protein: at the heart of the spliceosome*. RNA, 2005. **11**(5): p. 533-57.
 121. Gauthier, N., et al., *Cyclebase.org--a comprehensive multi-organism online database of cell-cycle experiments*. Nucleic acids research, 2008. **36**(Database issue): p. D854-D859.
 122. Gauthier, N., et al., *Cyclebase.org: version 2.0, an updated comprehensive, multi-species repository of cell cycle experiments and derived analysis results*. Nucleic acids research, 2010. **38**(Database issue): p. D699-D702.
 123. Ren, L., et al., *Analysis of the role of phosphorylation in fission yeast Cdc13p/cyclinB function*. J Biol Chem, 2005. **280**(15): p. 14591-6.
 124. Messier, V., D. Zenklusen, and S. Michnick, *A nutrient-responsive pathway that determines M phase timing through control of B-cyclin mRNA stability*. Cell, 2013. **153**(5): p. 1080-1093.
 125. Treck, T., et al., *Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast*. Cell, 2011. **147**(7): p. 1484-1497.
 126. Storchova, Z. and C. Kuffer, *The consequences of tetraploidy and aneuploidy*. Journal of cell science, 2008. **121**(23): p. 3859-3866.
 127. Heinrich, S., et al., *Determinants of robustness in spindle assembly checkpoint signalling*. Nat Cell Biol, 2013.
 128. Marguerat, S., et al., *Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells*. Cell, 2012. **151**(3): p. 671-83.

129. Peng, X., et al., *Identification of cell cycle-regulated genes in fission yeast*. Mol Biol Cell, 2005. **16**(3): p. 1026-42.
130. Jiang, Y., et al., *Involvement of transcription termination factor 2 in mitotic repression of transcription elongation*. Molecular cell, 2004. **14**(3): p. 375-385.
131. Larson, D., et al., *Real-time observation of transcription initiation and elongation on an endogenous yeast gene*. Science, 2011. **332**(6028): p. 475-478.
132. Mason, P. and K. Struhl, *Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo*. Molecular cell, 2005. **17**(6): p. 831-840.
133. Choubey, S., Sanchez, A., Kondev, J., *Deciphering transcriptional dynamics in vivo by counting nascent RNA molecules*. arXiv preprint arXiv:1311.0050, 2013.
134. Bertrand, E., et al., *Localization of ASH1 mRNA particles in living yeast*. Molecular cell, 1998. **2**(4): p. 437-445.

Annexe 1

Liste des sondes utilisées pour le smFISH dans *S. pombe*.

Organisme	Gene_Sonde	Type	Sequence
<i>S. pombe</i>	MDN1_1	Exon	ccactcaatcaacacatcca
<i>S. pombe</i>	MDN1_2	Exon	caaagtgttgcttggaacgt
<i>S. pombe</i>	MDN1_3	Exon	atthttgaccagggttggtg
<i>S. pombe</i>	MDN1_4	Exon	gtccgatgcatacaaaaacc
<i>S. pombe</i>	MDN1_5	Exon	gactttgcaggtttcacgaa
<i>S. pombe</i>	MDN1_6	Exon	ccgtaagtctttgtttctcc
<i>S. pombe</i>	MDN1_7	Exon	gagaccgtccaagaaacaaa
<i>S. pombe</i>	MDN1_8	Exon	ccctagtccaacaagaggta
<i>S. pombe</i>	MDN1_9	Exon	cggcacgtgagagttaaaaa
<i>S. pombe</i>	MDN1_10	Exon	ccaagtaaggagtatgcga
<i>S. pombe</i>	MDN1_11	Exon	cagccagaagaattggtctt
<i>S. pombe</i>	MDN1_12	Exon	cccaatttcgcagcaatttg
<i>S. pombe</i>	MDN1_13	Exon	ctgtggaatcgctaagatga
<i>S. pombe</i>	MDN1_14	Exon	ccaggttttggagaggata
<i>S. pombe</i>	MDN1_15	Exon	ttgtgtcaacacaccagggt
<i>S. pombe</i>	MDN1_16	Exon	gtgcgtgctcgatatttga
<i>S. pombe</i>	MDN1_17	Exon	aagacgctaagcacttcaga
<i>S. pombe</i>	MDN1_18	Exon	gcctttttcagaagagga
<i>S. pombe</i>	MDN1_19	Exon	ttaccacgtgaggaatta
<i>S. pombe</i>	MDN1_20	Exon	cgttgccaagaatctttgt
<i>S. pombe</i>	MDN1_21	Exon	cctcaacacattcatctggt
<i>S. pombe</i>	MDN1_22	Exon	gcaaggaacgatctttgtga
<i>S. pombe</i>	MDN1_23	Exon	cgcaaacacaatcgacgata
<i>S. pombe</i>	MDN1_24	Exon	gtgatataaaaacgcaaccc
<i>S. pombe</i>	MDN1_25	Exon	ggtcggaaaggatcataaat
<i>S. pombe</i>	MDN1_26	Exon	gcactctggagttcttgaga
<i>S. pombe</i>	MDN1_27	Exon	ccaacatatcgatagcctc
<i>S. pombe</i>	MDN1_28	Exon	caaaaccgtctctagaagga
<i>S. pombe</i>	MDN1_29	Exon	aactctatagcgacgttacg
<i>S. pombe</i>	MDN1_30	Exon	gtaaagcctttcagggtgt
<i>S. pombe</i>	MDN1_31	Exon	gccaagagatgaagatgtga
<i>S. pombe</i>	MDN1_32	Exon	ttgtttgaacacctgcagc
<i>S. pombe</i>	MDN1_33	Exon	caccaactagaagcaatggt
<i>S. pombe</i>	MDN1_34	Exon	tagtggcttaccagtacca

<i>S. pombe</i>	MDN1_35	Exon	caagccagctaaaagttgga
<i>S. pombe</i>	MDN1_36	Exon	gattccgtctgtgagacat
<i>S. pombe</i>	MDN1_37	Exon	ataaccgccaagcatatcag
<i>S. pombe</i>	MDN1_38	Exon	aagtgtggaagcgttgatag
<i>S. pombe</i>	MDN1_39	Exon	acctctcatgcaatggaaga
<i>S. pombe</i>	MDN1_40	Exon	gacgttgatgccatggatat
<i>S. pombe</i>	MDN1_41	Exon	ttccaacgaaaccgtcgtgc
<i>S. pombe</i>	MDN1_42	Exon	ggaagaggttgttgaccatc
<i>S. pombe</i>	MDN1_43	Exon	ttagcccactgatttcaag
<i>S. pombe</i>	MDN1_44	Exon	catgaagccattggaacctc
<i>S. pombe</i>	MDN1_45	Exon	ataccagtgccggaacga
<i>S. pombe</i>	MDN1_46	Exon	gcaaattaatctcgtctag
<i>S. pombe</i>	MDN1_47	Exon	atgagcttaaaagttggcca
<i>S. pombe</i>	MDN1_48	Exon	cactgagaagtataaccgat
<i>S. pombe</i>	MDN1_49	Exon	attggagtatatcgccacg
<i>S. pombe</i>	MDN1_50	Exon	acgtttccaacatctgtag
<i>S. pombe</i>	MDN1_51	Exon	atcgaaaagatggctcaagt
<i>S. pombe</i>	MDN1_52	Exon	accaatacaagagaaccga
<i>S. pombe</i>	MDN1_53	Exon	cctctcggataacatgttca
<i>S. pombe</i>	MDN1_54	Exon	gcaactgatatagttcagc
<i>S. pombe</i>	MDN1_55	Exon	accatcgagagaaagactct
<i>S. pombe</i>	MDN1_56	Exon	gaccagctccatcaacaaga
<i>S. pombe</i>	MDN1_57	Exon	cgaacggtatagtgtgtct
<i>S. pombe</i>	MDN1_58	Exon	cgacaaagttcggctcaatg
<i>S. pombe</i>	MDN1_59	Exon	ttggagcaatttccgttaca
<i>S. pombe</i>	MDN1_60	Exon	aaagggatcttctcaatcca

Organisme	Gene_Sonde	Type	Sequence
<i>S. pombe</i>	SPP42_1	Exon	gaggattccccggtgtaac
<i>S. pombe</i>	SPP42_2	Exon	ctgtgaaggaggctcaaatc
<i>S. pombe</i>	SPP42_3	Exon	tcttcacataacctggaggt
<i>S. pombe</i>	SPP42_4	Exon	tgctggagtcttgtttttc
<i>S. pombe</i>	SPP42_5	Exon	tttgcgtcctttcgtaa
<i>S. pombe</i>	SPP42_6	Exon	ttgattttgagaagcgcgc
<i>S. pombe</i>	SPP42_7	Exon	tgtttgaacataacctgct
<i>S. pombe</i>	SPP42_8	Exon	caggcggagatcagcttc
<i>S. pombe</i>	SPP42_9	Exon	ttcataatttgcgcaggtg
<i>S. pombe</i>	SPP42_10	Exon	agcacgaaacttctagaag
<i>S. pombe</i>	SPP42_11	Exon	cgcccaaataggaacgttta

<i>S. pombe</i>	SPP42_12	Exon	gcatgtggcaagtacttcaa
<i>S. pombe</i>	SPP42_13	Exon	tggecatgttctcaagcaatt
<i>S. pombe</i>	SPP42_14	Exon	tccctgtattcttccaagg
<i>S. pombe</i>	SPP42_15	Exon	agcaccggttacatgatata
<i>S. pombe</i>	SPP42_16	Exon	ggactctcgtttacaaaagt
<i>S. pombe</i>	SPP42_17	Exon	aaaatgcgggtcaattacgc
<i>S. pombe</i>	SPP42_18	Exon	acatcgttccccattgagca
<i>S. pombe</i>	SPP42_19	Exon	ttttcgatcacgcttctcg
<i>S. pombe</i>	SPP42_20	Exon	ggaaaccgcaaacgcttaaa
<i>S. pombe</i>	SPP42_21	Exon	tggecttctcatcgtcaaaag
<i>S. pombe</i>	SPP42_22	Exon	aatagcctctaacggctcta
<i>S. pombe</i>	SPP42_23	Exon	aaccaatccataacaggagc
<i>S. pombe</i>	SPP42_24	Exon	ggagtatcttcaaagcctt
<i>S. pombe</i>	SPP42_25	Exon	gataggttgaccatttaca
<i>S. pombe</i>	SPP42_26	Exon	ggtaaattaagctccaacg
<i>S. pombe</i>	SPP42_27	Exon	cactcaaaagttgtaacct
<i>S. pombe</i>	SPP42_28	Exon	tgggatagccacgttcaaag
<i>S. pombe</i>	SPP42_29	Exon	aaggtcaaatctcgacct
<i>S. pombe</i>	SPP42_30	Exon	tccaatctccattctggg
<i>S. pombe</i>	SPP42_31	Exon	ggaaacgctatccggtattc
<i>S. pombe</i>	SPP42_32	Exon	acgcttcgagctctagaatt
<i>S. pombe</i>	SPP42_33	Exon	cggttggtggtattcactta
<i>S. pombe</i>	SPP42_34	Exon	ctcaggaggaacgaaaacat
<i>S. pombe</i>	SPP42_35	Exon	aaaaaagccggaagatcagg
<i>S. pombe</i>	SPP42_36	Exon	ggggtttattatgggatccc
<i>S. pombe</i>	SPP42_37	Exon	ttcatgcaaagtgagttggc
<i>S. pombe</i>	SPP42_38	Exon	ttctatggcagaatcttcgg
<i>S. pombe</i>	SPP42_39	Exon	gcactggctgtgtgctcaaa
<i>S. pombe</i>	SPP42_40	Exon	gggtgcccaataaaaatca
<i>S. pombe</i>	SPP42_41	Exon	ccagatcgttattaaaagg
<i>S. pombe</i>	SPP42_42	Exon	catcttgcgcacgcttagtc
<i>S. pombe</i>	SPP42_43	Exon	taccaatgctttatcagtgg
<i>S. pombe</i>	SPP42_44	Exon	gattaggaggacaatgctct
<i>S. pombe</i>	SPP42_45	Exon	gacacacgaacttaacagg
<i>S. pombe</i>	SPP42_46	Exon	cgatttgggatgcgcatgt
<i>S. pombe</i>	SPP42_47	Exon	ctaagtaagctcgattggt
<i>S. pombe</i>	SPP42_48	Exon	accaatcgateagaagtact
<i>S. pombe</i>	SPP42_49	Exon	acaaacttgagaccagctt
<i>S. pombe</i>	SPP42_50	Exon	aattgcagcatgtgtacc
<i>S. pombe</i>	SPP42_51	Exon	tgtcaaccctttctgtgaa
<i>S. pombe</i>	SPP42_52	Exon	agtaaggcgaagtatttcgc

<i>S. pombe</i>	SPP42_53	Exon	actgtacatgagaatcgaca
<i>S. pombe</i>	SPP42_54	Exon	atgctcaatgtttcccaaa
<i>S. pombe</i>	SPP42_55	Exon	tgtaatccatcagctagctg
<i>S. pombe</i>	SPP42_56	Exon	taacgatacattcccgtaa
<i>S. pombe</i>	SPP42_57	Exon	ctgccgatcaatcgatatt
<i>S. pombe</i>	SPP42_58	Exon	gcttgaaatctttgcaagca
<i>S. pombe</i>	SPP42_59	Exon	gaccagtgtggaagcgataa
<i>S. pombe</i>	SPP42_60	Exon	acaacctggacctttgcca

Organisme	Gene_Sonde	Type	Sequence
<i>S. pombe</i>	POL1_1	Exon	ttatccccgctttctctt
<i>S. pombe</i>	POL1_2	Exon	tcagcaaatctttgagacgg
<i>S. pombe</i>	POL1_3	Exon	ttttccggcttgacgaaga
<i>S. pombe</i>	POL1_4	Exon	ttgcttctgatgtgctagc
<i>S. pombe</i>	POL1_5	Exon	cttgcatattcttctctg
<i>S. pombe</i>	POL1_6	Exon	aaagtcacctcatccagtc
<i>S. pombe</i>	POL1_7	Exon	attatccacataaccagccc
<i>S. pombe</i>	POL1_8	Exon	gactttgatcccactcatca
<i>S. pombe</i>	POL1_9	Exon	tcatcatcctcgtcactgta
<i>S. pombe</i>	POL1_10	Exon	gaccagacgatccttttca
<i>S. pombe</i>	POL1_11	Exon	cagctagaacgttcttcgat
<i>S. pombe</i>	POL1_12	Exon	gcgccaatttgttgattagg
<i>S. pombe</i>	POL1_13	Exon	caggttgatagcatttgct
<i>S. pombe</i>	POL1_14	Exon	cagccatgaattcatcttc
<i>S. pombe</i>	POL1_15	Exon	tggtcgattgagccaagaat
<i>S. pombe</i>	POL1_16	Exon	tttttgtcgagaggcgctc
<i>S. pombe</i>	POL1_17	Exon	cgacgtttggcatttgatgt
<i>S. pombe</i>	POL1_18	Exon	ttgaggaagcagctttctga
<i>S. pombe</i>	POL1_19	Exon	gggatcagaagaacaacgt
<i>S. pombe</i>	POL1_20	Exon	ggtacagtttctgggatgta
<i>S. pombe</i>	POL1_21	Exon	aataggaagactgctagcag
<i>S. pombe</i>	POL1_22	Exon	ttcttctgtctcagtgca
<i>S. pombe</i>	POL1_23	Exon	ttcattggtgactgtcga
<i>S. pombe</i>	POL1_24	Exon	aatgggtgcatcagatccaa
<i>S. pombe</i>	POL1_25	Exon	ggcttcatggcaatagattc
<i>S. pombe</i>	POL1_26	Exon	gacctgatcatgactcatca
<i>S. pombe</i>	POL1_27	Exon	ttttgaaagc gatgcggtgt
<i>S. pombe</i>	POL1_28	Exon	gtgatagaggttctgagaa
<i>S. pombe</i>	POL1_29	Exon	gttccaagataaagggtgc

<i>S. pombe</i>	POL1_30	Exon	tcagaaaagcagagacgtc
<i>S. pombe</i>	POL1_31	Exon	cgaagcaggaacatctacat
<i>S. pombe</i>	POL1_32	Exon	ccatcctctcagtaacatc
<i>S. pombe</i>	POL1_33	Exon	catctcgggtataatccatcc
<i>S. pombe</i>	POL1_34	Exon	caaatagacacaatgagccg
<i>S. pombe</i>	POL1_35	Exon	ccttgagaaagcacgaaaca
<i>S. pombe</i>	POL1_36	Exon	cggggaaggaagtataaaga
<i>S. pombe</i>	POL1_37	Exon	cgctgcaatagaatcttcag
<i>S. pombe</i>	POL1_38	Exon	gctacttcgtcgtaaacatc
<i>S. pombe</i>	POL1_39	Exon	ccttgacacccttttagat
<i>S. pombe</i>	POL1_40	Exon	cttgagactcttgatttcc
<i>S. pombe</i>	POL1_41	Exon	acctctaagtaatcagcggg
<i>S. pombe</i>	POL1_42	Exon	caaatcagtaggcagagcag
<i>S. pombe</i>	POL1_43	Exon	cggtgagaaagaagaacca
<i>S. pombe</i>	POL1_44	Exon	aaggcgggtattagttccaaa
<i>S. pombe</i>	POL1_45	Exon	acttaagacaaactgctcga
<i>S. pombe</i>	POL1_46	Exon	attttcagccaacaaggacc
<i>S. pombe</i>	POL1_47	Exon	cggcaccaactagcattttt
<i>S. pombe</i>	POL1_48	Exon	aggtgacgaacatcctattt
<i>S. pombe</i>	POL1_49	Exon	ggcgttttcgaagtaattc
<i>S. pombe</i>	POL1_50	Exon	aagactcatcaccgtcattg
<i>S. pombe</i>	POL1_51	Exon	tgctctttgttaatcagcgt
<i>S. pombe</i>	POL1_52	Exon	cataacgacttcttgcttgt
<i>S. pombe</i>	POL1_53	Exon	ggcaaacccttttcaatgtc
<i>S. pombe</i>	POL1_54	Exon	atacgagggcgatgtcattg
<i>S. pombe</i>	POL1_55	Exon	aaggacgaatcaacgaaaag
<i>S. pombe</i>	POL1_56	Exon	ctggcgagctagcttttcaa
<i>S. pombe</i>	POL1_57	Exon	tgcccaaaaataaacatcagg
<i>S. pombe</i>	POL1_58	Exon	cttcttcgaagtcgaccaat
<i>S. pombe</i>	POL1_59	Exon	tattaaacgaacgtggccat
<i>S. pombe</i>	POL1_60	Exon	tctacgaactgttgagagct

Organisme	Gene_Sonde	Type	Sequence
<i>S. pombe</i>	CDC13_1	Exon	agttaaacgacgggtagtca
<i>S. pombe</i>	CDC13_2	Exon	tatttgccaataggtgctgg
<i>S. pombe</i>	CDC13_3	Exon	gattttcgtcattgttgccc
<i>S. pombe</i>	CDC13_4	Exon	cacgggcaatatggtttgaa
<i>S. pombe</i>	CDC13_5	Exon	ttctgaagagtgcaagagc
<i>S. pombe</i>	CDC13_6	Exon	tggtttcttgccatttact

<i>S. pombe</i>	CDC13_7	Exon	cttcttaggaacgttggtgg
<i>S. pombe</i>	CDC13_8	Exon	gaaacatcatccaacgcatg
<i>S. pombe</i>	CDC13_9	Exon	tggaacaccttctttgtgt
<i>S. pombe</i>	CDC13_10	Exon	acagaagcagtagtgtgtct
<i>S. pombe</i>	CDC13_11	Exon	ctcgagagcacgacgggtac
<i>S. pombe</i>	CDC13_12	Exon	ctgttcagggattatagac
<i>S. pombe</i>	CDC13_13	Exon	ttcttgaagcgggttcac
<i>S. pombe</i>	CDC13_14	Exon	aaaaacagaaggttggcgac
<i>S. pombe</i>	CDC13_15	Exon	ccatgggttgaacagagtg
<i>S. pombe</i>	CDC13_16	Exon	agttgcttgatccttatgga
<i>S. pombe</i>	CDC13_17	Exon	tatctttcgaacaacgcgt
<i>S. pombe</i>	CDC13_18	Exon	ctatcacgggtaagttggg
<i>S. pombe</i>	CDC13_19	Exon	ccaatcttgagattcgggac
<i>S. pombe</i>	CDC13_20	Exon	gaaaccataagagggtcagc
<i>S. pombe</i>	CDC13_21	Exon	gggcatagtttcaattcca
<i>S. pombe</i>	CDC13_22	Exon	tgacgatccatataagtagg
<i>S. pombe</i>	CDC13_23	Exon	gcatcttccatgcaagctct
<i>S. pombe</i>	CDC13_24	Exon	aaccaatcggttaagtattcc
<i>S. pombe</i>	CDC13_25	Exon	ggaaacgagaatgcacttca
<i>S. pombe</i>	CDC13_26	Exon	acggccaaaaacaatgtttc
<i>S. pombe</i>	CDC13_27	Exon	gagcatacccgaatgataa
<i>S. pombe</i>	CDC13_28	Exon	gcaatgccaaccaattgtag
<i>S. pombe</i>	CDC13_29	Exon	acttgctagcgatgaacaag
<i>S. pombe</i>	CDC13_30	Exon	tttggactgaaggcacat
<i>S. pombe</i>	CDC13_31	Exon	cggcttgaagaatttctct
<i>S. pombe</i>	CDC13_32	Exon	taggacacgcaaatgtagc
<i>S. pombe</i>	CDC13_33	Exon	tcgggtaagcaagattaaac
<i>S. pombe</i>	CDC13_34	Exon	cgacgtaggaaattcatagg
<i>S. pombe</i>	CDC13_35	Exon	tcgtaaaagtcggcttttga
<i>S. pombe</i>	CDC13_36	Exon	atagcggcagcacactgttg
<i>S. pombe</i>	CDC13_37	Exon	caagcatttcctggcaagg
<i>S. pombe</i>	CDC13_38	Exon	attacgattccacggtccac
<i>S. pombe</i>	CDC13_39	Exon	cataacctgagtaatgcaca
<i>S. pombe</i>	CDC13_40	Exon	catgttgaacaggcttttgt
<i>S. pombe</i>	CDC13_41	Exon	cttgatccagtcacgaacaa
<i>S. pombe</i>	CDC13_42	Exon	cgccaagagggatagaattc