mRNPs biogenesis perturbation by the bacterial Rho factor : genomic analysis of the recrutment of the THO complex and its subunits

Valentin Beauvais

UNIVERSITÉ D'ORLÉANS

ÉCOLE DOCTORALE SANTE, SCIENCES BIOLOGIQUES ET CHIMIE DU VIVANT

Centre de Biophysique Moléculaire, UPR4301 du CNRS

THÈSE présentée par :

Valentin BEAUVAIS

soutenue le : **05 Juillet 2023**

pour obtenir le grade de : **Docteur de l'Université d'Orléans**

Discipline/ Spécialité : Biologie Moléculaire et cellulaire, Analyse Bio-Informatique

Perturbation de la biogénèse des mRNPs par le facteur bactérien Rho : analyse génomique du recrutement du complexe THO et de ses sous-unités

THÈSE dirigée par :

Dr. A. Rachid RAHMOUNI	Directeur de recherche, CNRS Orléans
Pr. Alain LEGRAND	Professeur des universités, Université d'Orléans
Dr. Igor STUPAREVIĆ	Professeur associé, Université de Zagreb

RAPPORTEURS :

Dr. Benoit PALANCADE	Directeur de recherche, CNRS Paris
Dr. Maxime WERY	Ingénieur de recherche, Institut Curie, Paris

---

JURY :

Dr. Bertrand CASTAING	Président du Jury, Directeur de recherche, CNRS Orléans
Dr. Benoit PALANCADE	Directeur de recherche, CNRS Paris
Dr. Maxime WERY	Ingénieur de recherche, Institut Curie, Paris
Dr. Antoine ROLLAND	Maitre de conférences, Université de Rennes
Pr. Alain LEGRAND	Professeur des universités, Université d'Orléans
Dr. Igor STUPAREVIĆ	Professeur associé, Université de Zagreb

INVITÉ :

Dr Kévin MOREAU	Chargé de recherche, Institut Curie, Bures-sur-Yvette

Remerciements :

Je souhaite commencer mes remerciements avec quelques lignes pour le Dr Rachid Rahmouni, qui nous a brutalement quittés le 14 avril 2021. Rachid fut le superviseur et le concepteur de ce projet de thèse, mais pas uniquement. Il fut également un mentor d'exception et ce à bien des égards. C'est grâce à Rachid que j'ai réussi à trouver ma voie dans la biologie par la bio-informatique durant mon stage de deuxième année de master. C'est par l'autonomie qu'il m'a accordée pendant ma première moitié de thèse que j'ai pu approfondir mes connaissances et mes techniques, en apprenant de mes erreurs. Pour finir, il m'a transmis la curiosité et la maturité scientifique d'un chercheur. Cette thèse n'est pas uniquement le résultat de quatre années de recherche et d'épanouissement scientifique, mais également mon hommage personnel à mon mentor.

I briefly switch to English to thank Dr. Igor Stuparevic for accepting to co-supervise the end of my thesis and overseeing the writing of the THO paper. Thank you for the patience you had to read my first manuscript and for the opportunity to visit your lab in Zagreb. I also thank all the members of the Laboratory of Biochemistry of Zagreb for their warm welcome during my stay. A special thanks to Dr. Bojan Žunar for his work on the THO manuscript.

Je voudrais par la suite remercier le Dr. Alain Legrand pour avoir accepté de reprendre cette thèse et d'avoir pris le temps de relire mon manuscrit de thèse. Profitez bien de votre retraite méritée !

Je remercie sincèrement les membres du jury, les Dr. Benoit Palancade, Maxime Wery, Antoine Rolland et Bertrand Castaing d'avoir expertisé ce travail. Merci pour les conseils qui ont permis d'améliorer ce manuscrit et pour cette défense de thèse très stimulante scientifiquement.

Je souhaite ensuite remercier mes deux encadrants, puis comparses de l'équipe Rahmouni, Kévin et Aurélia. Si la distance nous a séparés, vous avez toujours continué à prendre le temps de répondre à mes questions. Les moments passés en stage, puis en thèse, resteront précieux. J'aimerais particulièrement remercier Kévin pour ma formation au RTFM et pour le travail préparatoire sur THO, Aurélia pour tout le travail fourni à la paillasse, notamment pour la réalisation des librairies de séquençages pour les différents projets, et également pour sa patience durant mon stage (et devant mes bonhommes bâtons). Un merci tout particulier à toi d'avoir été présente durant le moment le plus difficile de cette thèse.

J'aimerais également remercier toutes les personnes qui ont permis à cette thèse d'aboutir, malgré les circonstances. Pour cela, je remercie notamment les Dr. Marc Boudvillain

et Matthieu Réfrégiers pour leur soutien ainsi que les directrices de l'école doctorale, les Dr. Agnès Delmas et Hélène Benedetti. J'en profite pour remercier les équipes administrative et informatique et plus spécifiquement Laetitia, Christine et Pierre pour leurs coups de main !

Un merci également au Dr. Christine Mosrin-Huaman pour son implication dans le projet THO.

Un remerciement particulier pour le Dr. Lucile Mollet pour m'avoir intégré à son équipe et pour l'opportunité professionnelle et scientifique qu'elle m'a ouverte. Merci aussi à Chloé et Kossi pour leur accueil et leur bonne humeur.

Un remerciement également pour mes camarades de master devenus compagnons de thèse et amis, Elodie et Christophe, parce que cette thèse aurait été bien plus dure sans nos pauses cafés, nos repas, Biotechnocentre et nos discussions, parfois scientifiques.

Un merci également à mes amis, Erwan, Loic et Alex. Nos parties de JDR m'auront bien aidé à décompresser pendant ces 4 ans

Un merci est également de mise pour ma famille, mes parents, mon frère, ma sœur et mes grands-parents pour leur soutien et pour ces moments reposants loin d'Orléans.

Pour finir, je souhaite remercier Clémence. Changer de bureau fut finalement une des meilleures choses qui me soit arrivées pendant cette thèse.

Summary

1. Introduction

Transcription is the biological process leading to the production of RNA from a DNA template. This process is present among all living organisms, either prokaryotes or eukaryotes, with some functional variations between these domains. Since transcription is the key mechanism between our biological code, our DNA, and the physical actors of our metabolism, the proteins, this process has been thoroughly studied through the last 60 years. The first evidence of RNA polymerase activity was discovered in 1959 by Weiss and Gladstone in rat liver [1]. Research on the transcription process has continued since then. Three forms of polymerase were discovered in 1969 in eukaryotes [2]. Those three complexes were named PolI, II and III. Large panels of biochemical work were done to discover the respective role of these 3 forms. If all those complexes ensure transcription of DNA in RNA, their DNA substrate and produced RNA differ. Indeed, PolI has been shown to synthesize large ribosomal RNA (rRNA) precursors, which are mainly dedicated for messenger RNA (mRNA) translation and protein formation. PolII transcribes protein-coding genes into messenger RNA, but also a large panel of transcripts commonly referred to as non-coding RNA, like cryptic unstable transcripts (CUT) and stable uncharacterized transcripts (SUT). PolIII produces transfer RNA (tRNA), small rRNA and U6 small nucleolar RNA (snoRNA), among other short non-coding RNA (ncRNA). In prokaryotes, only one form of polymerase is in charge of producing all sorts of RNA [3-5]. I will separate this introduction into four parts, starting with the RNA molecule and following with the transcription mechanism and its actors. The third part will be dedicated to the mechanism underlying the detection and degradation of aberrant mRNA. I will finish this introduction with a presentation of the two main tools I used during my PhD, the Rho factor and bioinformatics. In the following part, I will catalog a few main classes of RNA that are currently well described.

1.1 Global presentation of an RNA

An RNA (ribonucleic acid) is a polymer of four different types of nucleotides. The nucleotides are arranged in a sequence that determines the characteristics of the RNA, from its encoded protein (mRNA), its survivability (polyA tail, GC content), to its

catalytic abilities (catalytic activity of some ncRNA). RNAs are generated from a DNA template. The DNA template, the genome, is distributed on components called chromosomes (16 in yeast, 46 in human, up to 4 chromosomes in prokaryotes [6]). Chromosomes harbor specific loci called genes, each coding for an RNA. Some genes may code for messenger RNA (coding genes) or non-coding RNA (non-coding genes). The genome by itself contains various features aside from genes like enhancers and repressors, which will positively or negatively regulate gene expression and RNA biogenesis. Other elements like transposable elements are mobile sequences, some resulting from millions of years of evolution, which also regulate gene expression. Finally, satellite sequences are repetitions of nucleotides more or less long, distributed throughout the genome or at some specific chromosomal location, such as centromeres. Other features exist, and more are still to be discovered, which will help us better understand gene expression regulation and genomic structure complexity.

As mentioned earlier, several types of RNA cohabit in eukaryotes. Many players involved in the biogenesis of one type of RNA overlap and may have different activities in the biogenesis of other types of RNA. The next part will be a succinct presentation of RNA subtypes that will be mentioned further in this manuscript.

1.1.1 mRNA

The main role of messenger RNA is to transport the genetic information contained in the genome to the ribosome in the cytosol to produce the backbone of life, proteins.

mRNA has three separate domains whose length varies. The main part is the open reading frame (ORF), which contains the RNA sequence, which will be translated into a protein. In correctly matured mRNA, an ORF contains exclusively exons, the sequence remaining on the RNA after the co/post-transcriptional process called splicing, which will be detailed later. The length of mRNA coding genes varies greatly and can reach more than 100000 base pairs (good examples are Homo sapiens titin, TTN, and dystrophin, DMD) [7, 8]. The ORF is flanked by two untranslated regions (UTR). The 5’ UTR starts at the 5’ extremity of the transcript until the first nucleotide of the ORF, while the 3’ UTR starts at the last STOP codon of the ORF and goes on until

the 3’ extremity of the transcript. As their names indicate, these regions will not be translated and will not influence the sequence of the resulting protein.

The average length of 5' UTRs is roughly constant between species. Their average size is between 100 and 200 nucleotides. The average length of 3' UTRs is much more variable. It ranges from about 200 nucleotides in yeast to 800 nucleotides in humans on average [9]. Depending on the gene, the length of both 5' and 3' UTRs varies from a dozen nucleotides to a few thousand [10]. Despite the length of some UTRs, it seems that the presence of a single nucleotide is enough to initiate the translation [11].

UTRs play a role in the modulation of the rate of translation [12], the subcellular localization [13] and the stability [14] of the mRNA. They also take part in the regulation of the export of the mRNA out of the nucleus. UTRs may also be linked to the post-modification of specific bases. Mutated UTRs may be linked to severe pathologies, underlining their major role in the regulation of gene expression [15]. Finally, the 3’ UTR acts in a length-dependent manner to trigger non-sense mediated decay (NMD) a translational-coupled mechanism that led to the degradation of aberrant transcripts [16, 17].

The effect of UTRs can be explained by three factors. First, nucleotide motifs located within these UTRs can be targeted by specific RNA-binding proteins. Second, sequence elements located in the UTRs can interact with specific complementary non-coding RNAs. The latter phenomenon has been shown to play a key role in gene regulation [18]. Finally, the repetition of some sequences may recruit proteins that will modulate the translation of the UTR’s mRNA. A good example is the binding of CUG-binding proteins to CUG repeats, which regulate alternative splicing, mRNA degradation, and translation [19].

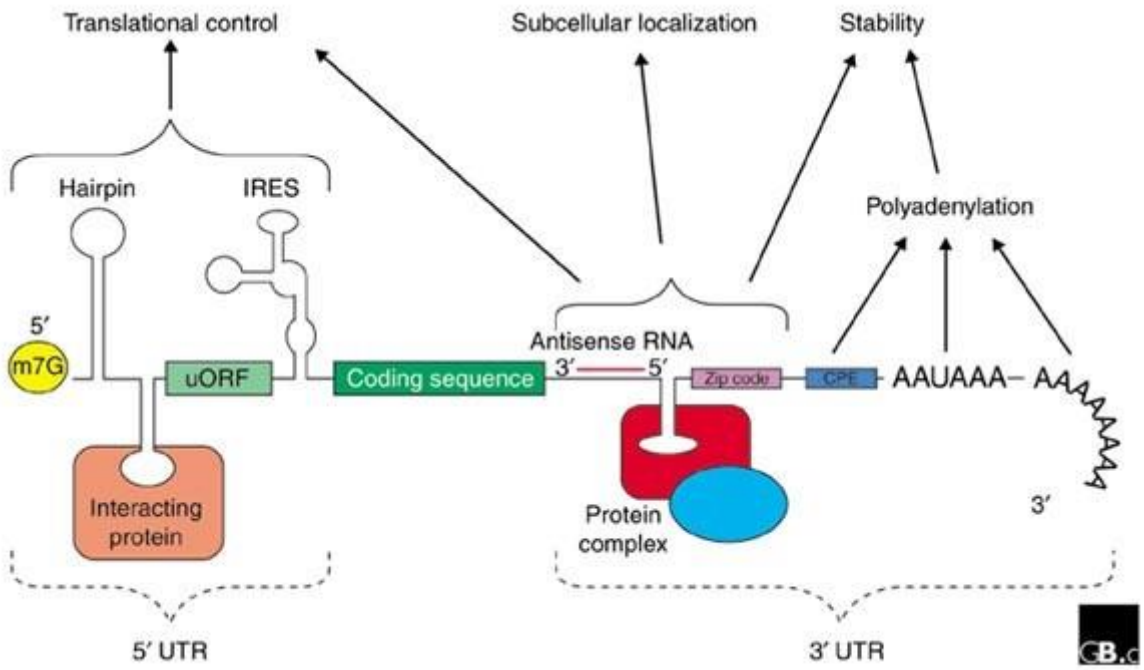


Figure 1: Composition of a typical mRNA, showing the implication of several post-transcriptionally added structures in the mRNA role and lifespan. Abbreviations (from 5' to 3'): UTR, untranslated region; m7G, 7-methyl-guanosine cap; hairpin, hairpin-like secondary structures; uORF, upstream open reading frame; IRES, internal ribosome entry site; CPE, cytoplasmic polyadenylation element; AAUAAA, polyadenylation signal. Taken from [9].

While the mRNA subfamily is crucial for protein coding, other subfamilies of RNA play vital roles. Non-coding RNA is a subfamily of RNA that, in contrast to mRNA, will not produce any protein, although recent studies points that peptides can arise from unexplored ncRNA translation [20]. They are divided into different classes depending on their role, their degradation rates and, for some, their sub-nuclear localization. Some ncRNAs were identified in more than one different subfamily due to separate teams discovering them with different techniques.

1.1.2 Long non coding RNA examples in yeast : Cryptic unstable transcripts (CUT), Xrn1-sensitive unstable transcripts (XUT) and stable uncharacterized transcripts (SUT)

This part will focus on specific subclasses of non-coding RNA called long non coding RNA (lncRNA). In yeast, they are mostly (but not only) divided in 3, non-exclusive categories : CUT, SUT and XUT.

lncRNAs also exists in human, subdivided in subclasses such as long intergenic non coding RNA (lincRNA), natural antisens transcript (NAT), enhancer RNA (eRNA), Promoter upstream transcript (PROMPT) and more. However, since my topic revolves around yeast, I will not explore these classes of transcripts in this manuscript.

CUT stands for cryptic unstable transcripts produced by polymerase II (PolII). This family of ncRNAs remains in the nucleus and is quickly degraded after or even during their transcription [21]. They are generated from intronic sequence, antisens loci and also through pervasive transcription [22]. Their roles are not very clear but some works indicate that they can negatively modulate the transcription rate of proximal genes via transcription interference, histone modification or DNA methylation [23]. The supercoiling from ncRNA transcription and the presence of the transcription actors could impair the opening of the chromatin and /or the recruitment of PolII to nearby genes [24]. The main hypothesis is that transcription of an ncRNA also blocks or impedes any transcription of the opposite stand, which may contain coding loci, effectively repressing the gene. Another hypothesis includes the regulation of transcriptions factors and regulators, acting as “storage templates” as long as they are not needed on more important loci. These RNAs undergo post-transcriptional modification before their degradation like capping and polyA tailing, though the last process is different between mRNA and CUT [25].

These transcripts are degraded by the exosome through a pathway detailed below. However, a sub-population of these transcripts is degraded by the exonuclease Xrn1 in the cytoplasm and is called XUT [26].

SUT stands for stable uncharacterized transcripts. Their role seems to be quite similar to that of CUT, but they are more stable as their degradation pathway is different. They are still poorly understood. Some studies indicate that their processing could be handled in a similar way than that of mRNA. [27].

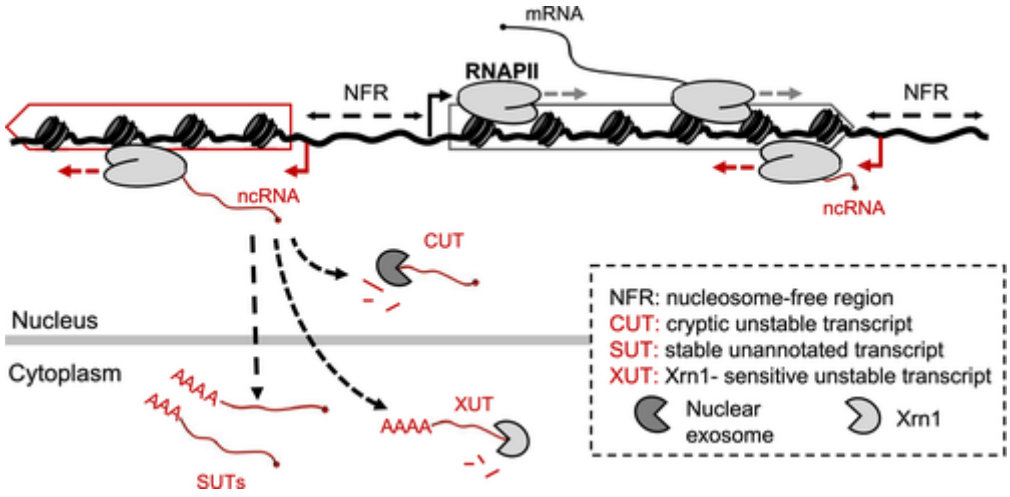


Figure 2: Overview of CUT/XUT and SUT biogenesis through pervasive transcription and degradation. Cryptic promoters are indicated by red arrows, whereas promoters of protein-coding genes are denoted by black arrows. Taken from [22]

1.1.3 tRNA

tRNA stands for transfer RNA. They are a special class of RNA as their main role occurs during translation, as they bring the amino-acid corresponding to a specific codon (sequences of 3 nucleotides present on the mRNA) to the ribosome active site. This amino acid is then incorporated into the nascent protein during translation. tRNAs are typically composed of 70 to 100 nucleotides in length in eukaryotes and are heavily modified during their maturation [28]. The structure of the tRNA is also conserved between different types as a cloverleaf shape, adapted to fit the tRNA through the A sites of a ribosome. The key site of a tRNA is the anticodon, a set of 3 nucleotides that will pair with the corresponding codon of an mRNA in the ribosome. The tRNA bore the amino acid corresponding to its anticodon at its 3' end [29].

1.1.4 rRNA

rRNA stands for ribosomal RNA. These rRNAs are the main and essential components of the ribosome, the massive machinery responsible for the translation of mRNA into proteins. rRNAs are separated in two subcategories: “small subunit” and “large subunit” depending on the ribosomal subunits they belong to. The small subunit monitors codon-anticodon base-pairing between the mRNA and tRNAs, while the large subunit, which harbors the catalytic peptidyltransferase activity, is responsible for the synthesis of the nascent polypeptide chain [30].

Type	Size	Large subunit (LSU rRNA)	Small subunit (SSU rRNA)
prokaryotic	70S	50S (5S: 120 nt, 23S: 2906 nt)	30S (16S: 1542 nt)
eukaryotic	80S	60S (5S: 121 nt, 5.8S: 156 nt, 28S: 5070 nt)	40S (18S: 1869 nt)

rRNA precursors have their bases post-transcriptionally modified (see figure below) and bent by folding proteins to ensure adequate structuration of the rRNA [31]. An association of snoRNA (small nucleolar RNA), and proteins exerts further endo and exo-nucleolysis modifications to the rRNA. Aside from sequence modifications, rRNA bases can also undergo modifications, detailed in Figure 3. The assembly of the various mature rRNA and ribosomal proteins leads to the formation of a mature ribosome.

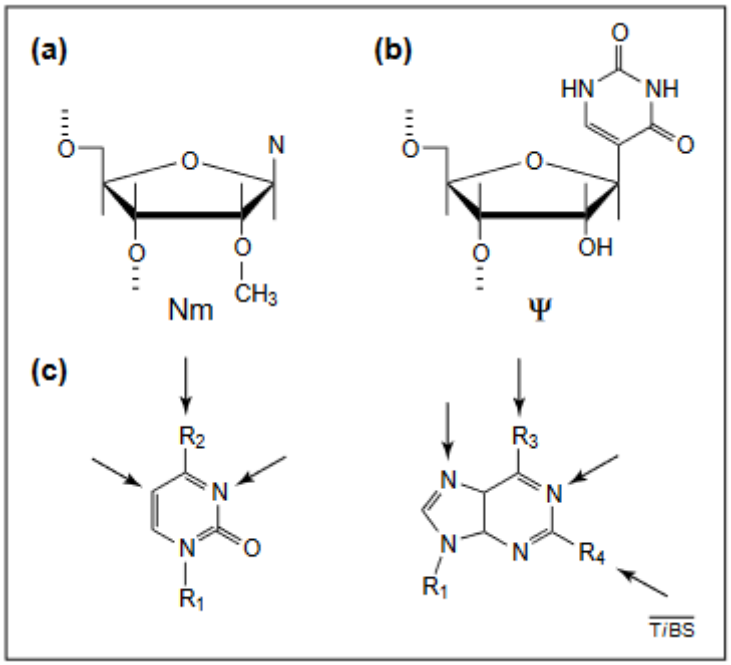


Figure 3: Major types of rRNA modification. (a) 2'-O-methylation (Nm). (b) Isomerization of uridine to pseudouridine (Ψ), 'the fifth nucleoside'. (c) Most other modifications involve base methylation at various positions (arrows). Taken from [31].

The degradation of rRNA uses a different pathway than mRNA and ncRNA, using the non-functional rRNA decay pathway, which will not be developed in this manuscript. The degradation of rRNA already included in a fully formed ribosome is linked with the ubiquitination of the latter. Some actors implicated in the degradation of other RNA types, such as the exosome, are implicated in the maturation of the 3' end of the 5.8S rRNA [32].

1.1.5 snoRNA

snoRNA stands for small nucleolar RNA. These transcripts are short (60–300 nt long), mostly nucleoli-localized, non-polyadenylated ncRNAs, present in all eukaryotic organisms. snoRNAs are mainly encoded by intronic regions of both protein coding and non-protein-coding genes [33]. These RNA products play a role in the maturation of rRNA precursors and snRNA (small nuclear RNA). They are also implicated in mRNA splicing and chromatin maintenance. SnoRNAs are divided into three classes, orienting their actions: C/D box snoRNAs and H/ACA box snoRNAs.

C/D box snoRNAs guide−2′-O-ribose methylation, and H/ACA box snoRNAs direct the pseudouridylation of nucleotides. A third type of snoRNA called scaRNA is located in Cajal bodies. They mostly function as the two other types but some of them have both sites [34]. snoRNAs associate with various proteins to ensure their function : Nop1p, Nop56p, Nop58p, and Snu13p for C/D box snoRNAs, Cbf5p, Gar1p, Nhp2p, and Nop10p for H/ACA box snoRNAs [35]. The snoRNAs target rRNAs but are also implicated in microRNA (miRNA) formation and snRNA/tRNA maturation [36, 37]. The maturation of the 3’ end snoRNA is handled by the exosome and the degradation of the aberrant ones is handled by the same actors implicated in CUT degradation [38].

1.1.6 snRNA

snRNAs stands for small nuclear RNAs (<150 nt) and are commonly found in nuclear compartments called splicing speckles. They play a critical role in the maturation of mRNA and more specifically in the splicing step. snRNAs interact with many proteins to form RNA-protein complexes, termed as small nuclear ribonucleoproteins (snRNPs), in the cell nucleus. Five small nuclear RNAs (snRNAs U1, U2, U4, U5, and U6) participate in pre-mRNA splicing by recognizing the critical sequence elements present in the introns, thereby forming active spliceosomes, along with a large number of splicing protein factors [39]. The recognition of a primary target site is achieved primarily through base-pairing interactions between the 10 highly conserved nucleotides at the 5' end of U1 snRNA and the intron sequences of the pre-mRNA at the 5' splice site (G/GUAUGU in yeast or G/GURAGU in vertebrates) or nucleotide-nucleotide contact between U1 snRNAs and pre-mRNA [40].

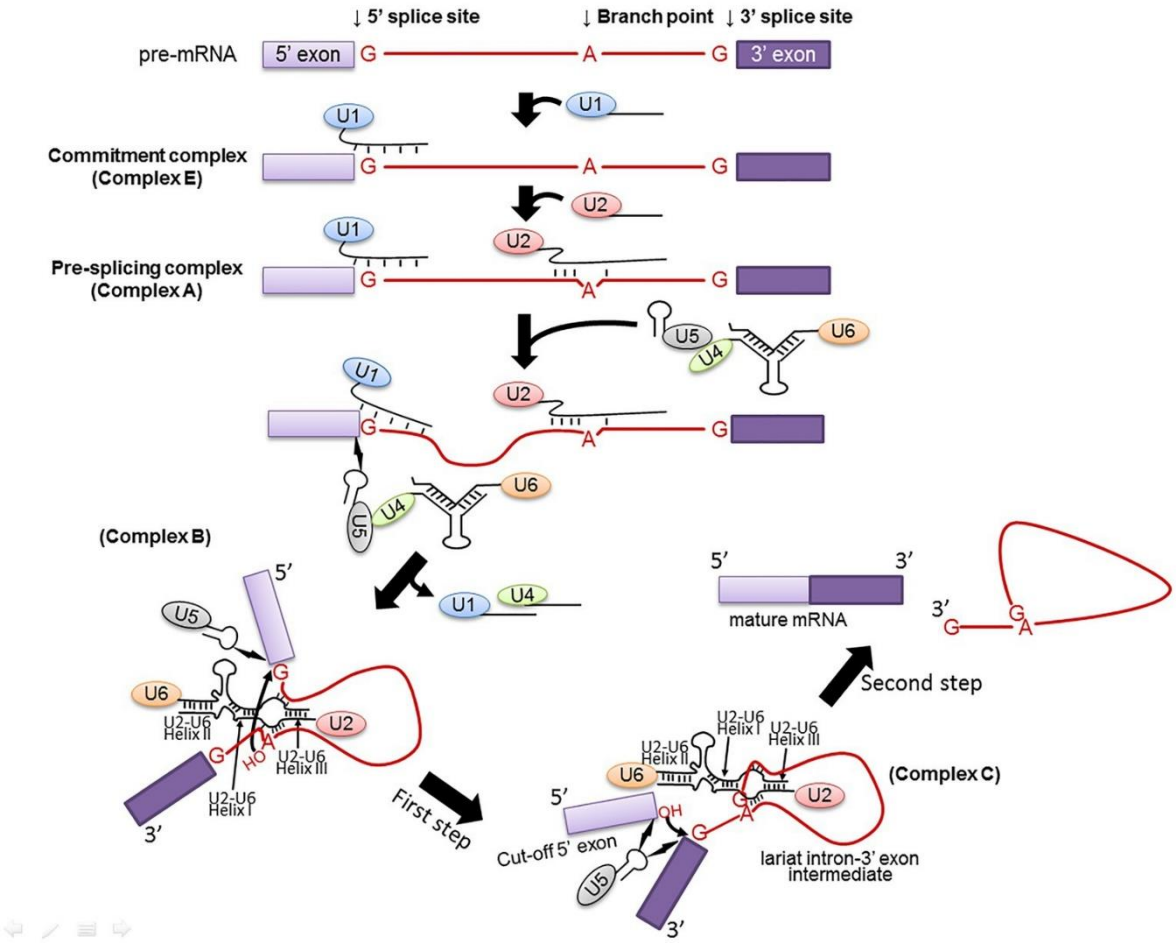


Figure 4: Spliceosome assembly. Spliceosome assembly is a dynamic multi-step process. Shown are several important steps resulting in Complexes E, A, B, and C. The branch site, the 5' and 3' splice sites are also shown. The 5' and 3' exons are in light and dark purple boxes, respectively, and the intron is in red line (and red letters). U1, U2, U4, U5, and U6 snRNPs are also depicted. RNA-RNA interactions, including U1–5' splice site, U2-branch site, U6–5' splice site, U4–U6, and U2–U6 (Helixes I, II, and III) are indicated as well. The two curved arrows indicate nucleophilic attacks (transesterification reactions). The lightning bolts indicate non-Watson-Crick nucleotide-nucleotide contacts. Taken from [40].

Notably, snRNAs are extensively modified with different RNA modifications, which confer unique properties to the RNAs. The most abundant modified nucleotides in snRNAs are pseudouridine (Ψ) and 2'-O-methyl residues, whereas m6A and m2G are rarely present in only a few snRNA species. These modifications are handled by the snoRNPs as mentioned before [36, 37, 40].

1.1.7 miRNA, siRNA and piRNA

These three different classes of small non-coding RNA are all involved in gene silencing, although they do not exert this activity through the same pathway [41]. Although these 3 RNA classes do not exist in *S.Cerevisiae,* they are important for the regulation of gene expression in other species, notably in humans [42].

MicroRNA (miRNA) are a class of small non-coding RNA of 22 nucleotides in length in average. Their role is mostly regulatory as they are able to interact with the 3'UTR of mRNAs. The interaction of the miRNA with its partial or complete complementary sequence on the mRNA triggers its de-polyadenylation and de-capping before ensuring its degradation. In addition, this fixation also inhibits translation. [43, 44]. Other studies suggest alternative interaction sites in the coding sequence, the 5' UTR and even the promoter of genes [45]. In opposition to their usual silencing activity, the interaction of miRNAs with gene promoters can induce their transcription [46].

Small/short interfering RNAs (siRNAs) are small RNA between 20 and 25 nucleotides. Similarly to miRNA, they interact with mRNA through complementary base-pairing but can target any 25nt long sequence on the mRNA [44]. The silencing effect of the siRNAs is called RNA interference (RNAi). The siRNA, after maturation, is incorporated within the RISC complex (RNA induced silencing complex). There, it will serve as a targeting strand against a specific sequence of an mRNA. Upon fixation of the siRNA-RISC complex on its target, Ago2, a protein from RISC, cleaves the mRNA [41]. SiRNA and the RNAi system are well-known for their potential therapeutic applications. In fission yeast, siRNAs are also implicated in heterochromatin assembly via the RITS complex (RNA-induced initiation of transcriptional gene-silencing complex), which is distinct form RISC. The RITS complex cleaves nascent transcripts from a targeted gene before recruiting chromatin remodeling actors such as Clr-C [47, 48].

Piwi-interacting RNAs (piRNAs) are 21–35 nucleotides long. They silence transposable elements, regulate gene expression and fight viral infections. piRNAs guide PIWI proteins (P-element-induced wimpy testis proteins) to cleave target RNA,

promote heterochromatin assembly and methylate DNA [49]. However, the way they exert their function is still poorly understood.

In this manuscript, we focus on PolII, transcription and regulation of mRNA with a few words on ncRNA biogenesis. Their transcription can be divided into 3 parts: initiation, elongation and termination, which will be discussed in more detail further in this manuscript. The main actor, RNA polymerase II, also plays the role of conductor of the whole mechanism. It serves as a recruitment platform for all the cofactors needed for the proper production of RNA but also for the cofactors that will mature the RNA (i.e. make the RNA export-competent). My topic focuses on one aspect of RNA biogenesis and more particularly on the degradation of messenger ribonucleic proteins (mRNPs). An mRNP is the assembly of an mRNA and various proteins that will be detailed in this manuscript. The assembly of the mRNP is a necessary step in the life of an mRNA to fulfill its purpose.

1.2 PolII-mediated transcription

The transcription of mRNA and ncRNA by PolII involves not only the polymerase but also a variety of co-factors that play a role in the polymerization of the RNA strand. The production of RNA by PolII is subdivided into three steps: initiation, elongation and termination, each of which involves different co-factors. In addition, co- and post-transcriptional modifications occur to achieve maturation of the RNA. These modifications consist of an addition of nucleotides such as the 5’ cap or the polyA tail. It can also be a deletion of a portion of the sequence by a phenomenon called splicing. Modifications of some specific nucleotides can also occur, although they will not be explored in this manuscript. In addition to these modifications of the RNA strand, several proteins will coat the transcript, without modifying its sequence. Once all those modifications are applied, the mRNP is considered mature and export competent.

1.2.1 PolII

PolII is the main catalytic actor of the mRNA and ncRNA transcription since it handles the polymerization of the ribonucleic chain. In addition to this primordial activity, it also plays a conductor role, recruiting co-factors and maturing complexes when needed.

1.2.1.1 Structure

The core of the PolII is composed of 12 subunits (Rpb1 – 12), the larger one being Rpb1. Ten of them form the polymerase core, which associates with Rpb4 and Rpb7 to form the fully functional heterocomplex [50]. Thanks to structural techniques like X-ray crystallography, interesting structures were discovered and some hints on the functioning of PolII were discovered. First, the entrance to the active site can switch between an open and closed state by a clamp guiding the melted template, a single DNA strand, through an opening to the active site. The opening of the clamp is also impacted by the fixation of other proteins such as initiation or elongation factors, which may explain at least partially how these proteins work to influence transcription rate. After the entrance, near the active site, a trigger loop opens and closes onto each to-be-added NTP and detects base pair mismatches, effectively enforcing the fidelity of the PolII. Still near the active site, after base-addition, a wall separates the DNA and RNA strands. At this point, the DNA template makes a 90 degree turn to exit the PolII. Rpb1, the main subunit of PolII, works in coordination with the wall to successfully separate RNA from DNA and guides both RNA to an exit channel and the DNA to a positively charged protrusion which helps the DNA strands to re-hybridize [51, 52].

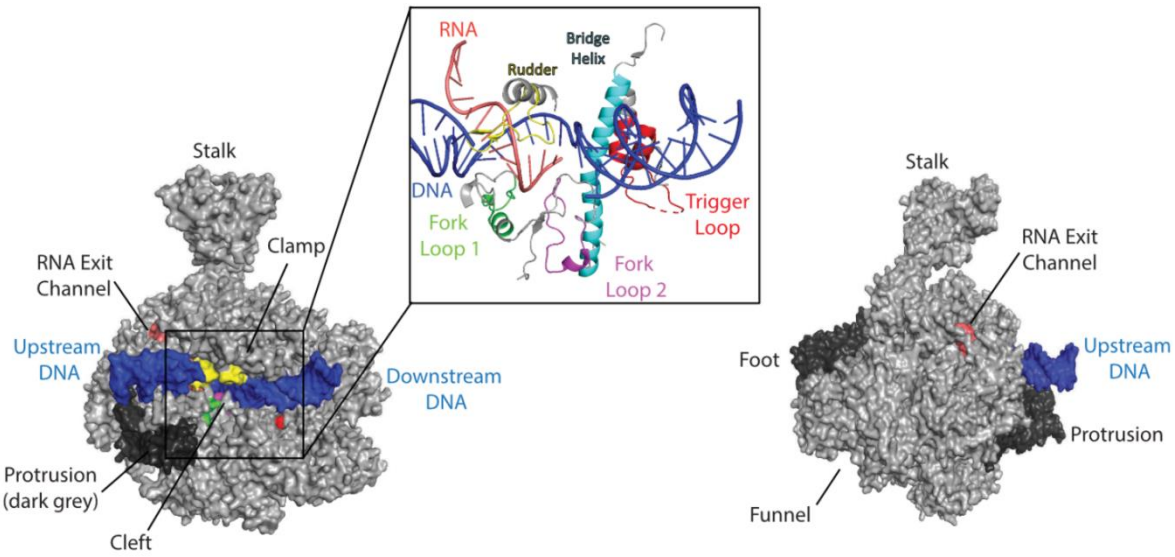


Figure 5: Overview of the PolII structure during transcription. Bovine Pol II (gray) shown in two orientations (rotated 180°). DNA is colored blue, and the Pol II stalk, clamp, foot, funnel, and RNA exit channel are marked; the protrusion and foot domains are shown in dark gray. Bovine PolII are considered identical to the human enzyme except for seven amino-acids and are largely use for structural studies [53]. Figure taken from [51]

1.2.1.2 CTD of Rpb1

The polymerase does not operate alone during the transcription but also works as a platform for other proteins, giving the entire complex abilities needed to perform the transcription. The recruitment of these factors is closely linked to the C-terminal domain (CTD) of the polymerase main subunit, Rpb1.

The polymerase CTD is mostly unstructured and composed of 26 repetitions of the Tyrosine-Serine-Proline-Threonine-Serine-Proline-Serine peptides in yeast (52 in humans). [51]. The repeats may be phosphorylated or glycosylated on Tyr1, Ser2, Thr4, Ser5, Ser7 and the prolines can be isomerized. This CTD can undergo various modifications, which are responsible for different patterns of recruitment of the PolII co-factors at each stage of the transcription [54].

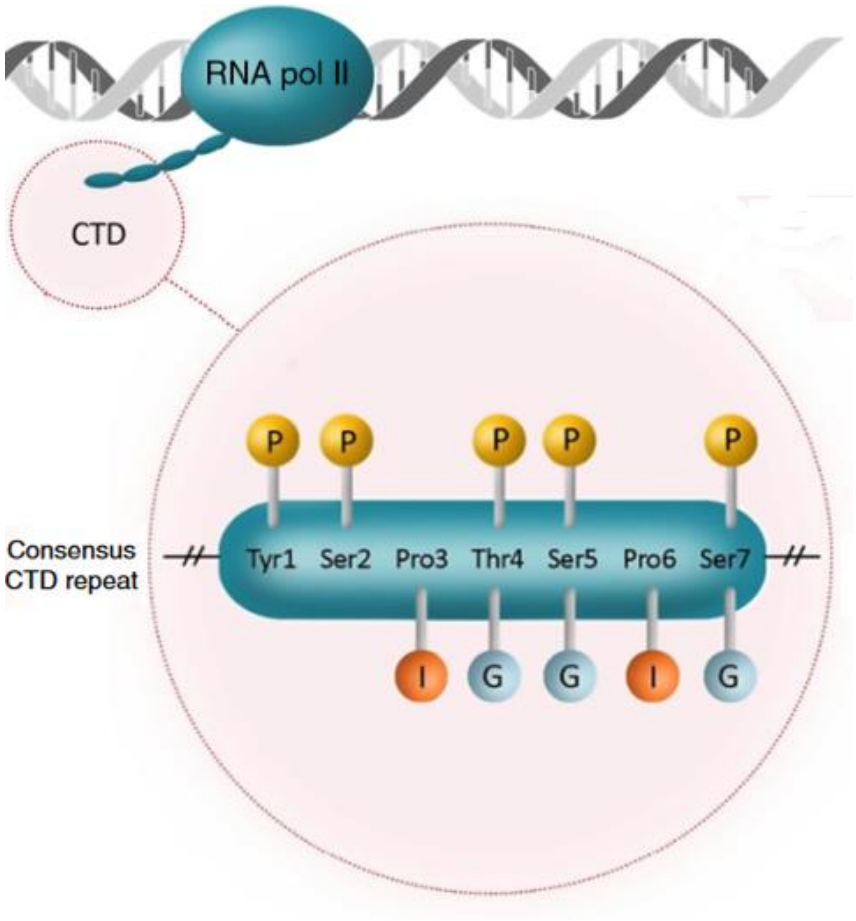


Figure 6: Residues of the PolII consensus CTD repeats. The figure shows the respective modifications each type of residue can undergo. Yellow circles (P), phosphorylation; blue circles (G), glycosylation; orange circles (I), isomerization. Taken from [54].

The patterns of post-translational modifications of the CTD-repetitions change throughout the transcription cycle of mRNA coding genes. The most studied modifications are the phosphorylation of Ser2 and Ser5. In yeast, ChIP analysis revealed that Ser5 is highly phosphorylated (Ser5P) near the transcription start site (TSS) while Ser2 phosphorylated form (Ser2P) is mostly present at the 3' end of the genes. mNET-seq analysis also showed that Ser5P is also present at the 3' end of exons. Ser2P and Ser5P represent 75% of the total phospho-counts. Tyr1P is distributed along the gene body and is less present near the 3' and 5' ends. Thr4P is distributed uniformly along the genes with a peak near polyA sites. The patterns of modifications of both residues are different in humans. Modifications on heptads are generally present on one residue at a time although phosphorylation of both Ser2 and Ser5 can occur on rare occasions. [55]

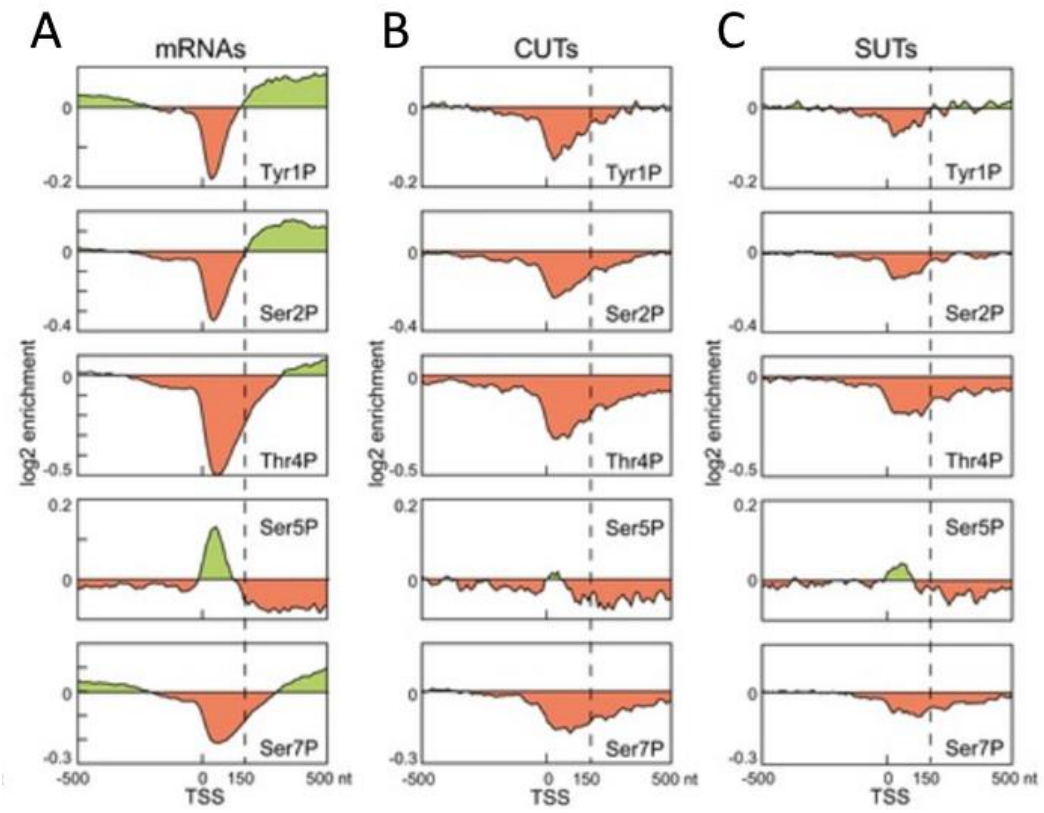


Figure 7: Evolution of the phosphorylation states of Tyr1, Ser2, Thp4, Ser5 and Ser7 around the TSS of mRNA/CUT/SUT coding genes. A) *Metagene analysis of RNAPII phosphorylation enrichment relative to transcription start sites, calculated for all mRNA genes. The TSS-proximal 150 nt region, where Ser5P is enriched and Ser2P and Tyr1P are depleted, is indicated with a dashed line. B) Metagene analysis of RNAPII phosphorylation on CUTs as for (A). C) Metagene analysis of RNAPII phosphorylation on SUTs as for (A). Taken from [55]*

These modifications are applied by specific enzymes as summarized below [54]. These enzymes can modify other proteins and are suspected to act on both the CTD and other transcription actors.

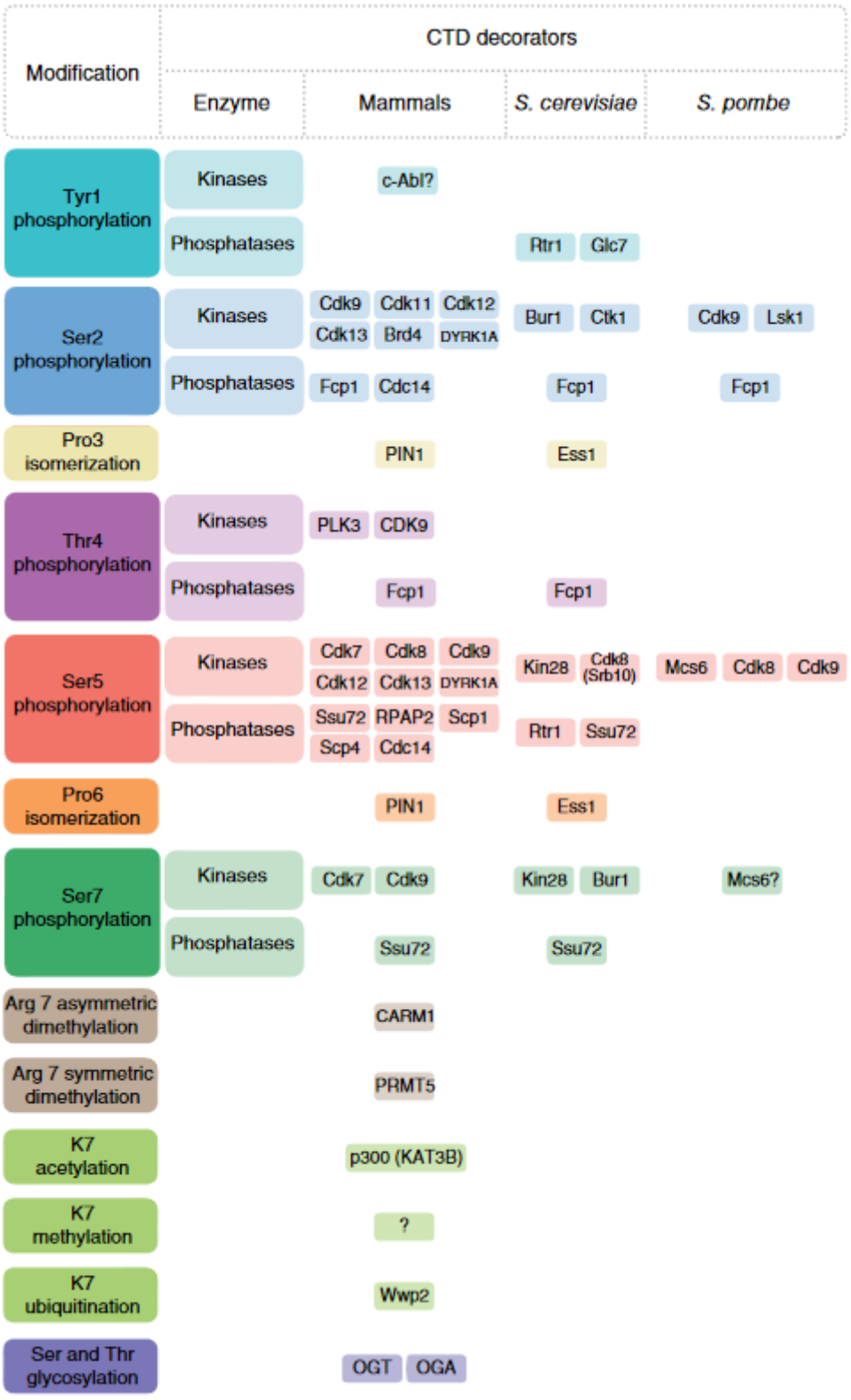


Figure 8: Actors contributing to the post-translational modification of the CTD residues in Mammals, S.cerevisiae and S.pombe. [54]

Both Ser2 and Ser5 modifications induce recruitment mRNA processing factors like the CAP-adding enzymes by Ser5-P, the THO complex during elongation and Pcf11 near the transcription end site (TES) by Ser2-P. The modulation of the phosphorylation of the CTD is also linked to histone dynamics, through the recruitment of methyl/acetyl-transferases, thus remodeling the chromatin and promoting transcription [56]. In mammals, the phosphorylation of Ser5-P near the 3' ends of exons is correlated with splicing, highlighting a link between transcription and the splicing machinery [57]. However, the actors recruited by the modifications of other residues are still unclear.

Recent studies and reviews explain that the CTD of PolII is implicated in condensates inside the nucleus. These condensates occur thanks to the interaction between low-complexity domains (similar to the PolII CTD) from histone ubiquitination factors or splicing factors. The PolII CTD phosphorylation state links PolII with these condensates and supports a transition between "initiation" and "elongation" condensates [58, 59].

In addition, it has been shown that some actors implicated in the cell cycle like cyclin-dependent kinases (CDKs) compete with PolII-CTD phosphatase to condition the phosphorylation state of the CTD. This contributes to the modulation of transcription activity throughout the cell cycle [60].

1.2.2. Transcription steps and co-factors

In conjunction with the main core, transcription factors are a key to the transcription process. As mentioned above, some of them are directly recruited by the PolII core, possibly via the CTD of Rpb1. Their recruitment is sequential and step-dependent.

1.2.2.1 Initiation

Initiation represents the first step of the transcription of messenger RNAs. For clarity, I will detail here the most common aspect of transcription initiation related to core promoters and I will overlook the mechanisms linked to alternative transcription start site. PolII assembles with TFIIB, TFIID, TFIIE, TFIIF, and TFIIH on a promoter DNA sequence to form the pre-initiation complex (PIC). The steps leading to the PIC assembly (transcription preinitiation complex) may vary depending on the transcribed

gene, but order-of-addition experiments made it possible to develop a classic model of how it usually occurs. In this classical model, TFIID, consisting of TBP (TATA-binding protein) and TAF (TBP associated factors), interacts with a TATA box-promoter leading to the bending of the promoter DNA locus. However, ~80 % of functional promoters are TATA-less in eukaryotes [61]. It is admitted that TAFs are able to target short promoting sequences, effectively anchoring TFIID on TATA-less promoters [61]. Other sequence elements associated with the promoter such as BRE (TFIIB recognition elements) are used by other subunits to anchor the DNA [62]. TBP then recruits TFIIA and TFIIB. Pol II and the TFIIF complex interact with a DNA-bound TFIIB factor, forming the core of the PIC. The role of TFIIF is to prevent non-specific interaction of PolII with DNA and influence TSS selection (transcription start site). It also stabilizes the PIC. The functional completion of PIC requires the supplemental fixation of both TFIIE and TFIIH, as these cofactors induce the promoter DNA opening. TFIID is a non-mandatory component of PIC. However, it was shown that it contributes to promoter recognition and is implicated in the expression of the most actively PolII-transcribed gene. In the presence of free nucleotides, the complete PIC opens the constitutive DNA of the promoter, forming a transcription bubble. In this bubble, the DNA template strand is encompassed in the PolII active site, which triggers the RNA polymerization from the DNA strand. [63]

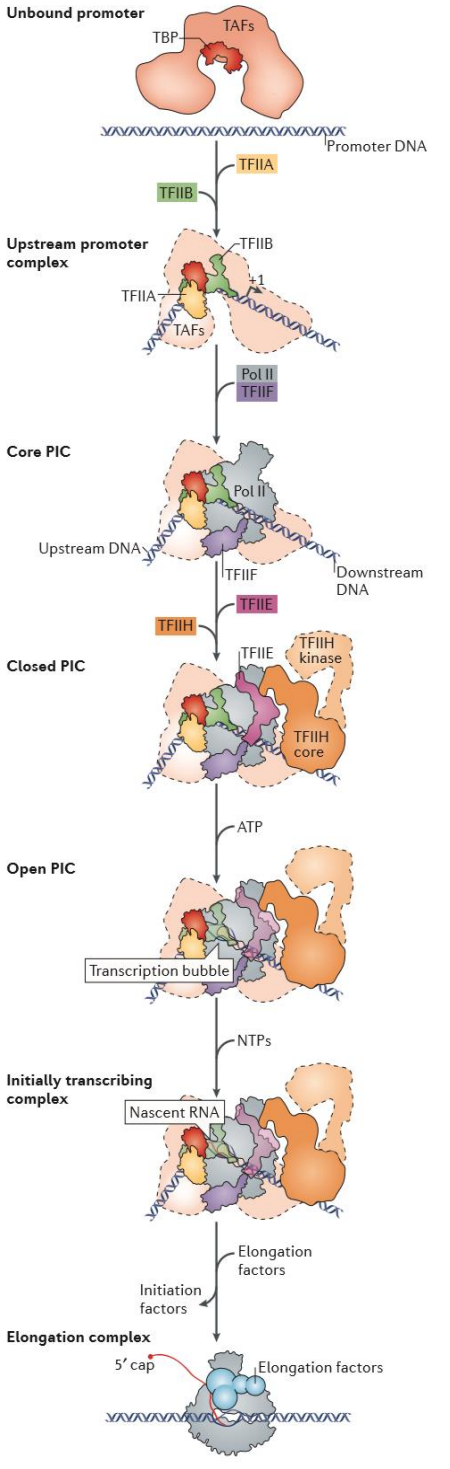


Figure 9: Step-by-step description of the co-factor recruitment and conformational change of the Polymerase II complex from the pre-initiation to the elongation step. The names for the

intermediate complexes that form during the initiation–elongation transition are provided to the left of the images. Taken from [63].

The PIC is quite unstable in this state and, in many cases, leads to a series of aborted transcripts before the PolII is able to reach productive elongation after its release from the promoter. *In vitro*, some of the PIC subunits remain on the promoter as a scaffold complex, which can then be used to easily launch a new round of transcription [64].

1.2.2.2 Elongation

The initial steps of elongation occur after the transcription of a few dozens of nucleotides downstream of the transcription start site (TSS). At this point, the PolII marks a pause. This phenomenon called “pausing” is linked to transcription regulation and also to a quality checkpoint to implement the 5’ capping of the transcript and perform PolII modification before productive elongation. It has been shown that this pausing depends on the environment (sequences, distance from TSS, recruited proteins) of the core promoter recruiting PolII in the first place, since it modulates the ability of PolII to detach from the promoter [65]. TFIIH is implicated in the release of pausing due to its promoter DNA melting properties leading to the release of PolII [66]. The pausing is maintained jointly by the DRB-sensitivity inducing factor (DSIF) and the negative elongation factor (NELF) [67]. PolII escapes from the promoter-proximal region through the mediation of the positive transcription elongation factor (P-TEFb) complex. It is recruited to the PolII and phosphorylates its CTD at Ser2, as well as NELF, evicting it, but also DSIF, switching it from a negative to a positive regulator. This leads to the release of PolII from the promoter proximal pause site and PIC dissociation, thus marking the start of the productive elongation phase [68]. The nucleosomes may also contribute to pausing by competing with PolII for chromatin occupancy. A recent study suggested that in yeast, 2 pauses could occur at the start of the elongation, the first due to the release from the promoter and the second to the removal of the +1 nucleosome [65].

Elongation rates vary between genes and are linked to co-transcriptional processes from splicing to transcription termination, histone marks, cleavage and polyadenylation sites. DNA sequence composition also affects elongation rate, as low-

complexity sequences with low CG content have a higher elongation rate [68]. PolII can also pause due to backtracking on certain DNA sequences during elongation. In this case, the transcription is resumed by TFIIS, which triggers the intrinsic cleavage activity of PolII, removing over transcribed RNA and allowing the transcription to resume from the re-aligned active site [69]. Aside from this mechanism, other transcription factors are involved in resuming transcription after pausing. However, their mechanisms of action are still unclear [70]. PolII pauses after the transcription of cleavage and polyadenylation sites before transcription termination and its subsequent release [71].

1.2.2.3 Termination

Transcription termination depends on numerous factors, especially the type of transcribed RNA. As many termination factors interact with degradation pathways, termination is also a regulatory step in gene expression. The various termination systems have been thoroughly studied in the last decade specifically in yeast [72]. mRNA coding genes transcription is terminated via the CPF-CF pathway. Components of this pathway are recruited to the RNA either via the 3' untranslated region (UTR) of the nascent transcript or via the PolII CTD [73]. Upon recruitment of all required proteins, the nascent RNA is cleaved and a polyA tail is added to the nascent transcript by a process that will be detailed later in this manuscript. Correctly processed mRNPs are then exported to the cytoplasm and their half-life depends on the cytoplasmic RNA decay pathway. The polymerase then transcribes downstream DNA for a length rarely exceeding 150nt, before being evicted.

Historically, there are two models to explain the eviction of PolII. The first one, called the torpedo model, is mediated by the 5'-3' exonuclease Rat1 (Xrn1 in human), which targets the unprotected 5' end of the residual RNA still protruding out of the PolII after the cleavage of the RNA. Rat1 degrades the remaining RNA until it reaches PolII, leading to the dissociation of the polymerase from the chromatin. The mechanism that leads to this dissociation is still unclear, as it could be linked to the torsion of RNA or the steric cluttering from Rat1. The second model explains that the binding of the termination complex on the nascent transcript leads to a conformational change of the elongation complex provoking the eviction of elongation promoting factors [72].

A hybrid-model combining both allosteric and the torpedo model was proposed more than a decade ago in yeast [74]. In humans, a unified model taking characteristics from the two previous models has been described [75]. The first step of this model is the deceleration of PolII transcription past the polyadenylation site (PAS) orchestrated by dephosphorylation of the elongation factor SPT5. In human, the phosphorylation of the PolII CTD on the Thr4 could also be implicated either in an imminent termination signal or an allosteric change, as this modification is prominent after the PAS [76, 77]. In any case, the resulting slowness of PolII allows Rat1/Xrn2 5'-3'exonuclease to catch up quickly with the enzyme after RNA cleavage [78].

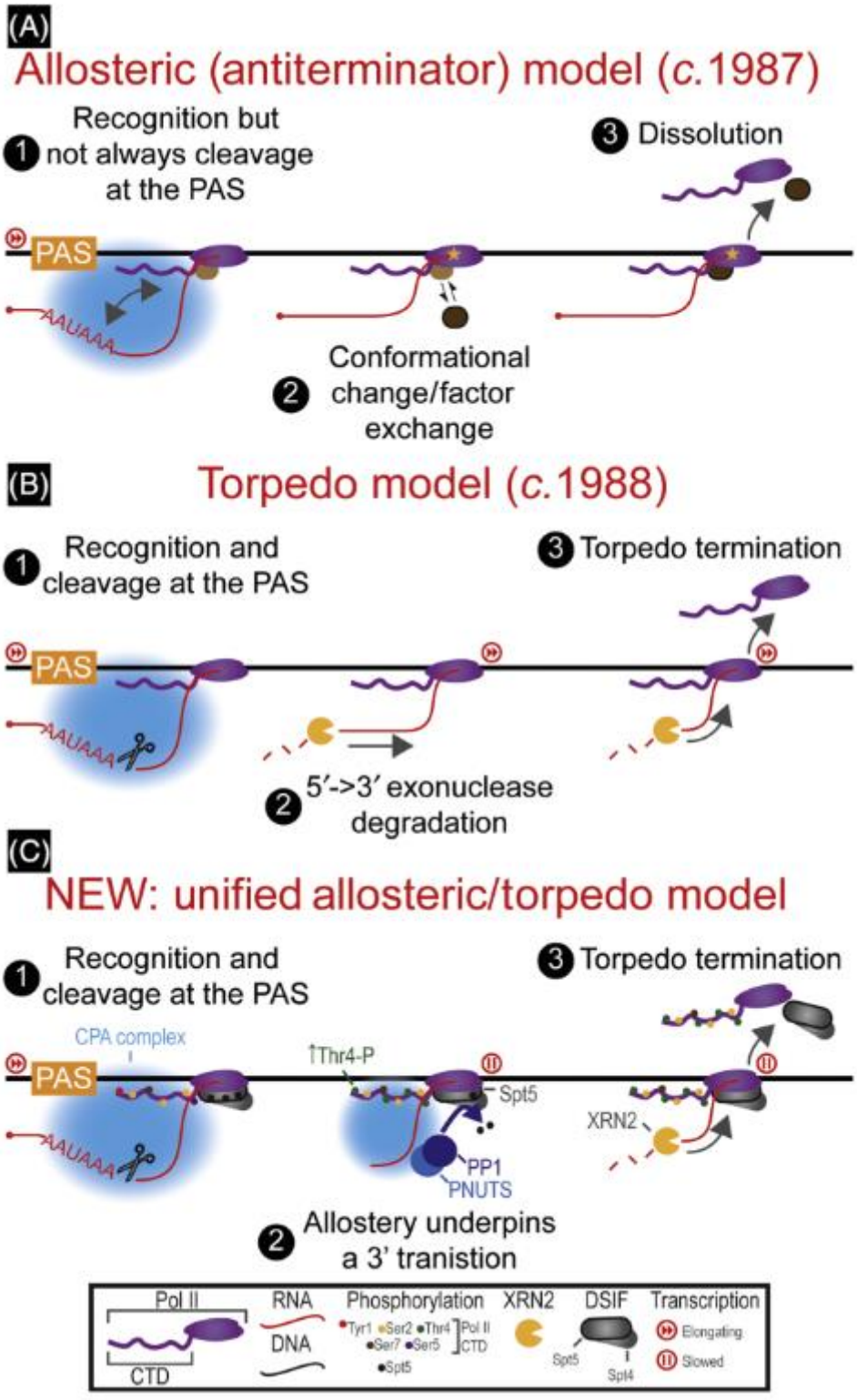


Figure 10: Summary of the two exclusive historical models and the more recent, unified model for PAS-Dependent Pol II Termination. Taken from [77].

The transcription termination of ncRNA in nucleus is mainly managed by the NNS-dependent pathway. In contrast to the mRNAs that travel to the cytoplasm to be translated, lncRNAs, and more specifically CUTs, are processed and degraded shortly after their transcription. The NNS is composed of 2 RNA interacting proteins (Nab3 and Nrd1) and one helicase (Sen1). Nab3 and Nrd1 recognize specific motives on the nascent transcripts and are dropped from PolII CTD onto the RNA [79, 80]. Subsequent fixation of both proteins leads to the recruitment of Sen1, therefore completing the complex. RNA-interacting Nrd1 then recruits a second complex, TRAMP, composed of the polyadenylate polymerase Trf4/5, the zinc-knuckle protein Air1/2 and the helicase Mtr4 . This complex cleaves the nascent transcript and also adds a polyA tail on the 3' end. Direct recruitment of the exosome by TRAMP leads to the degradation of the targeted transcript [80]. The actors implicated in the degradation of the ncRNA will be explored further below.

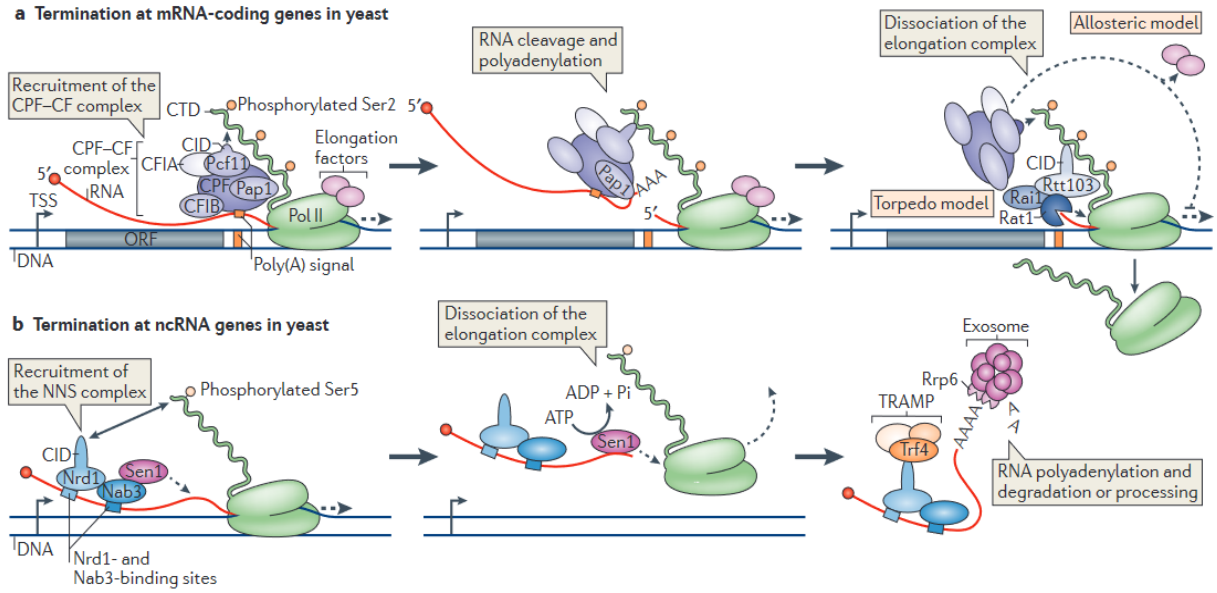


Figure 11: Termination of the transcription of mRNA and ncRNA. (a) Termination at mRNA-coding genes in yeast. The CPF-CF factor is composed of CPF, CFIA and CFIB. In addition, Pcf11 interacts with Ser2-phosphorylated form of the CTD of the PolII through its CTD-interaction domain (CID). Pap1 is a CPF-associated polymerase. (b) Termination at non-coding genes. The different actors present in this illustration will be detailed in this manuscript. Taken from [72]

1.2.3 mRNP formation and mRNA maturation

To be functional, a nascent mRNA undergoes multiple modifications. These modifications can be made directly on the RNA, like the 5' cap or the splicing of introns, or by the addition of proteins that will have specific roles in transcript protection, quality control or nuclear export for example. The association between an mRNA and its packaging proteins is called a messenger ribonucleoprotein (mRNP)

1.2.3.1 CAP

The first maturation that occurs to the transcript is the addition of a protective structure at its 5' end. The addition of a 5' cap occurs very early during transcription. Fundamentally, the cap is a 5'-5' methylguanosine, which protects the transcript from any 5'-3' exonuclease activity [81].

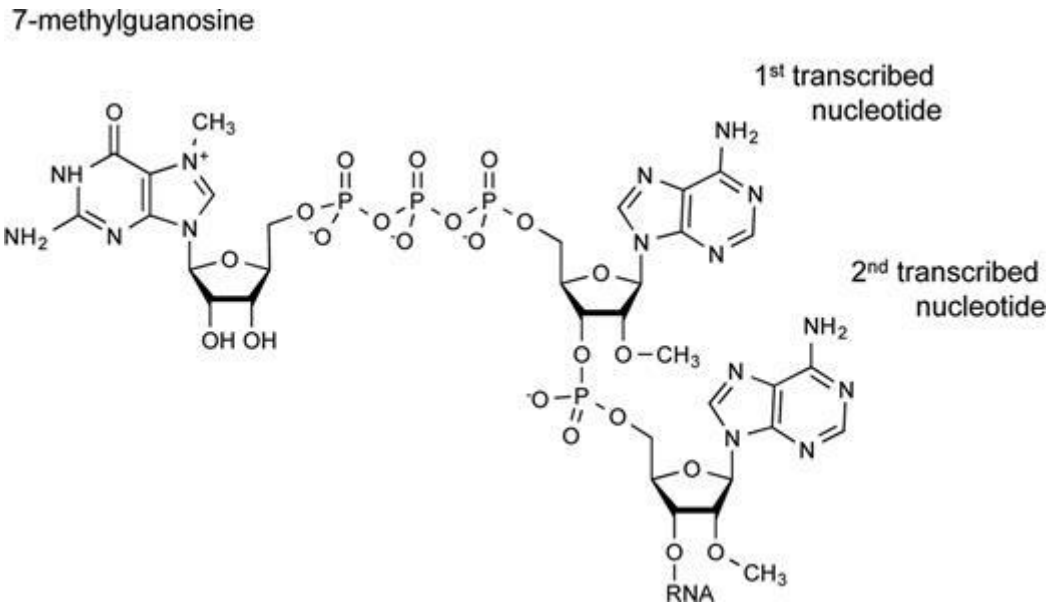


Figure 12: Structure of the 5' cap. Taken from [81]

The addition of this cap occurs during pausing, after transcription initiation but before productive elongation. Three enzymes are recruited onto the nascent transcript by the phosphorylated Ser5 of the PolII CTD. The first one is a RNA triphosphatase, which will remove the gamma-P linked to the base at the 5' end of the transcript. Then, an RNA guanylyltransferase adds a guanosine-tri-phosphate to the 5' end, releasing a pyrophosphate as a byproduct. Finally, the cap is completed by the addition of a methyl to the nitrogen in position 7 of the guanosine by an RNA-methyltransferase. To

complete the structure, a protein complex then binds to the 5'-5' methylguanosine. The two most documented complexes are the elf4F and the CBC (cap-binding-complex). The latter is composed of two subunits, Cbp20 (20kDa) and Cbp80 (80kDa). This complex adds supplementary protection against exonuclease activity. Furthermore, the presence of a protein complex is needed for specific translational events [82]. In the traditional model for mRNA translation, the CBC is generally removed from the mRNA after successful export to the cytoplasm and is replaced by elF4E via the action of importins. However, eIF4E can also work as a functional nuclear cap-binding factor on its own and the CBC is linked to translation via the CBC-dependent translation factor (CTIF) [83]. A review deemed these two pathways functionally interchangeable, adding that the CBC pathway is preferred for the translation of a subset of mRNA and for stress-dependent responses [84].

1.2.3.2 Splicing

In eukaryotes, some transcripts are spliced from some parts of their sequence during elongation. In budding yeast, the genes coding spliced transcripts represent a limited portion of the entire mRNA-coding gene population (~ 6%) but they are highly transcribed [85]. In humans however, this mechanism is far more generalized [86]. Splicing consists of the elimination of a portion of the transcribed RNA called introns. The mature RNA resulting from the splicing step contains non-intronic sequences called exons. The splicing mechanism is complex but is summarized in the figure below.

Although traditional splicing of mRNA removes introns and keeps exons, alternative splicing is a form of splicing that regulates the exons present in the final mRNA product. This mechanism allows the same gene to produce two or more different mRNAs and, subsequently, different proteins. Alternative splicing is one of the main explanations for the transcriptome diversity in eukaryotes [87, 88].

The relationship between PolII and spliceosome recruitment is currently described by two non-exclusive models. The first one relies on the ability of the PolII CTD to recruit the RNA binding proteins (RBPs) to the RNA, including the spliceosome subunits. In the same way as for cap-adding enzymes, the CTD should be phosphorylated on its Ser5 but also on its Ser2 to be able to recruit the spliceosome.

The second model is called kinetic coupling and consists of the availability of competing splicing sites depending on transcription speed. A higher transcription rate could lead to more splicing sites being accessible, thus inducing competition between them, leading to exon-exclusion [89].

1.2.3.3 Packaging by proteins: the THO complex

In addition to the capping during transcription initiation and splicing during elongation, protein cofactors interact with and remain on the mRNA, effectively packaging it. Their importance to the fate of the mRNA is diverse, ranging from structural protection to export capabilities. One example of these cofactors is the THO complex.

1.2.3.3.1 Structure

Described two decades ago in yeast, the THO complex (THOc) is a key factor in mRNP biogenesis [90]. It is composed of four main proteins, Tho2 (180 kDa), Hpr1 (90 kDa), Mft1 (45 kDa) and Thp2 (30 kDa). Tex1 (47 kDa), even if it is not critical for complex formation and integrity, is often considered as a fifth, non-essential subunit. Cryo-EM analysis has shown that THO dimerizes upon in-vitro reconstruction. The dimerization seems to be mediated by Thp2 and Mft1 subunits [91]. Immunodepletion by anti-Tho2 antibodies leads to complex dislocation in high salt concentration medium [92]. The regulation of THO complex formation could be carried out via Hpr1, as it can be ubiquitinated by the ubiquitin ligase Rsp5 pathway [93].

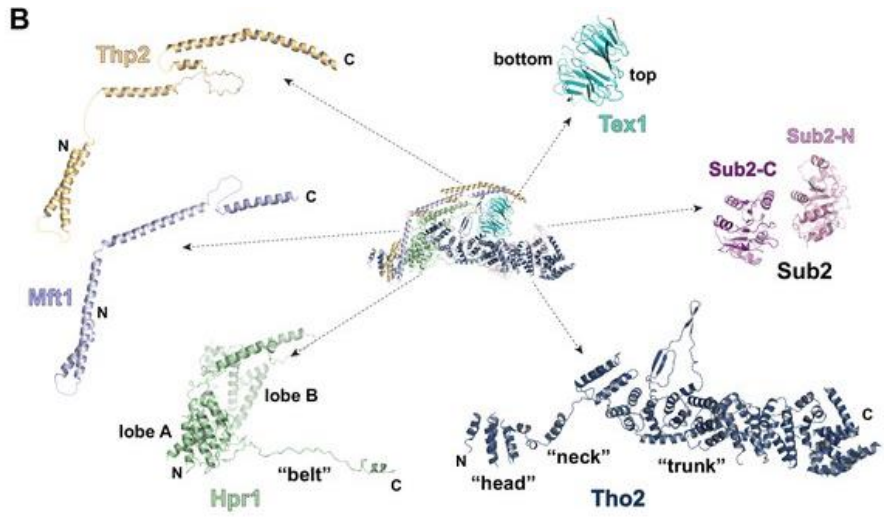


Figure 13: THO complex subunits. Tex1 is considered as a fifth, non-critical subunit. Taken from [94]

This complex then associates with the SR-like proteins Sub2 (UAP56 in human) and Yra1 (Ref/Aly in human) to form the TREX complex. The THO-Sub2 complex is still a dimer but it is supposed that only one single copy of Yra1 interacts with the THO-Sub2 dimer. The interaction of THO with Tex1 and Sub2 is mediated by Tho2 and Hpr1. [95]. The roles of TREX will be explored below.

The interaction of THO with its RNA substrate is handled by Tho2. The Tho2 C-terminal extremity is an unfolded tail mostly composed of highly positively charged residues. Multiple sequence alignments did not reveal any conserved or identified sub-regions or motifs within Tho2-CTD but its truncation led to a loss of Tho2-RNA interaction. Although not critical to THO complex stability, the loss of Tho2-CTD leads to impairment of gene expression similar to the phenotypes observed upon full Tho2 depletion and impedes with the interaction of THO with active chromatin [96].

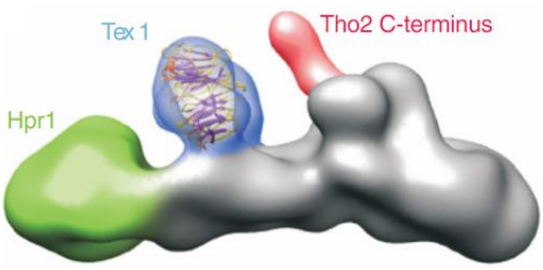


Figure 14: The Tho2 C-terminus is placed on a protrusion atop THO. Taken from [96].

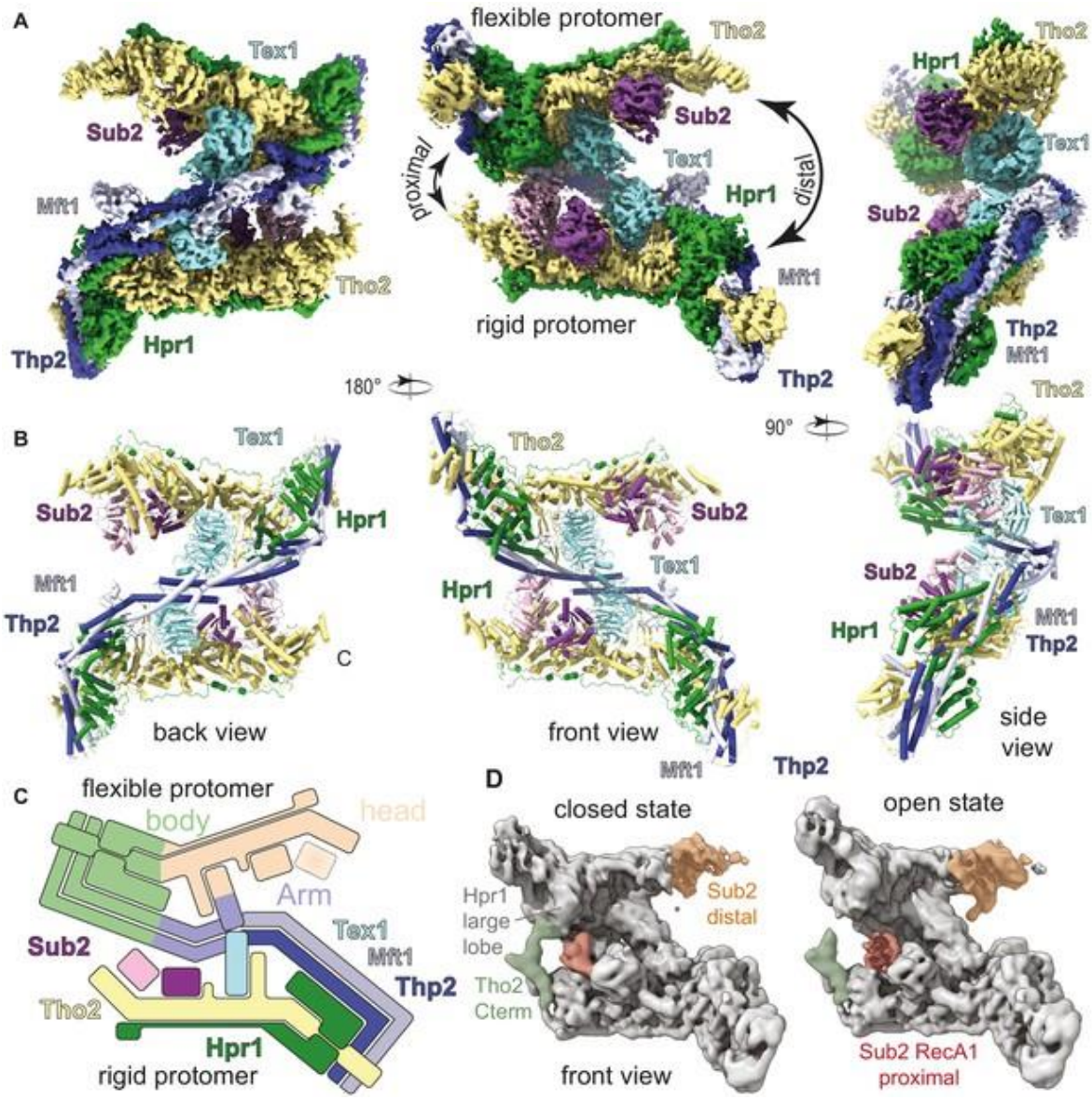


Figure 15: Organization of the THO-Sub2 complex dimers. (A) Segmented cryo-EM reconstruction of the THO-Sub2 dimer. (B) Cartoon representation of the structure, shown in the same orientations and colors. Helices are rendered as solid cylinders. (C) Schematic representation of the THO-Sub2 complex architecture based on the cryo-EM structure. (D) Two frames of the raw cryo-EM data outputs from the variance analysis. The Tho2 C-termini of the two protomers are shown in orange and green. [91]

Two SR-like proteins are also known to interact with THO while also playing a role in splicing: Gbp2 and Hrb1. They are known for shuttling spliced mRNP to the nuclear membrane and Gbp2 interacts with the C-terminal domain of Tho2 for its loading on the transcript [94, 97]. In the case where splicing fails, Gbp2 and Hrb1

recruit the TRAMP complex, a component of the mRNP quality control (QC) process, leading to transcript degradation [98].

1.2.3.3.2 Recruitment and role

The THO complex (THO) is co-transcriptionally recruited to the chromatin by interacting with PolII CTD to the nascent transcripts [99]. The recruitment of THO on the RNA-PolII depends on the phosphorylation state of the CTD. In yeast, Ser2 needs to be phosphorylated to ensure THO recruitment to the chromatin. Ctk1 could also be a co-recruiter of THO, as a genetic interaction with Mft1 has been demonstrated [97]. Since Ser2 from the PolII-CTD is progressively phosphorylated during elongation, THO is more and more recruited onto the PolII and subsequently on the nascent transcript as the elongation progresses [100].

The association of THO with Yra1 and Sub2 is required for the export of the mRNA, since Yra1 is recruited after Sub2 action and serves as an adaptor for the mRNA export factor Mex67/Mtr2 in yeast, NXF1/NXT1 (TAP•p15) in humans [101]. The complex Mex67/Mtr2 interacts with the nuclear pore and allows the export of the mRNP.

Aside from this key role in mRNA export and processing, THO plays an important role in R-loop prevention. R-loops are a common phenomenon occurring during transcription where the DNA from the transcription bubble interacts with the nascent RNA provoking transcription failure and genetic instability. It is unclear whether the R-loops stalls the PolII from which the RNA moiety is born or if the R-loops blocks upstream PolII [102]. Thp2 is at least partially implicated in R-loops resolution, especially on telomeric loci [103].

Furthermore, the depletion of one of the THO-Sub2 complex subunits leads to the accumulation of aggregates composed of nucleic acids and 3’ poly-adenylating/cleaving proteins in heat shock conditions. This implicates that the THO complex also plays a role in the stability of mRNA [104-106].

Finally, THO is implicated in the polyA tailing process. The deletion of THO components leads to inefficient polyadenylation as Fip1 levels, a protein implicated in the cleavage and polyadenylation complex, are decreased [107].

1.2.3.4 PolyA tail and export

The last co-transcriptional modification of the mRNA is the addition of a poly-adenosine (polyA) tail.

The canonical polyadenylation steps for mRNA start with the transcription of a polyA signal, generally an AAUAAA or a close variant, followed by a GU/U rich region. The transcription of this site leads to the recruitment of the CPF/CPSF (cleavage and polyadenylation factor in yeast / cleavage and polyadenylation specificity factor in human), both on the pA sites and on the C-terminal domain of the PolII [73, 108]. This factor handles the cleavage of the mRNA, removing any excessively transcribed nucleotide, and adding a chain of adenosine. Besides protecting the transcript from 3'-5' exonuclease activity, the PolyA-tail and associated proteins are a requirement for the export of mRNA. In yeast, Nab2 is the reference protein linking the PolyA tail and export. This protein interacts with PolyA+ transcript and with Mex67, a key protein allowing the export of a transcript to the cytoplasm [109].

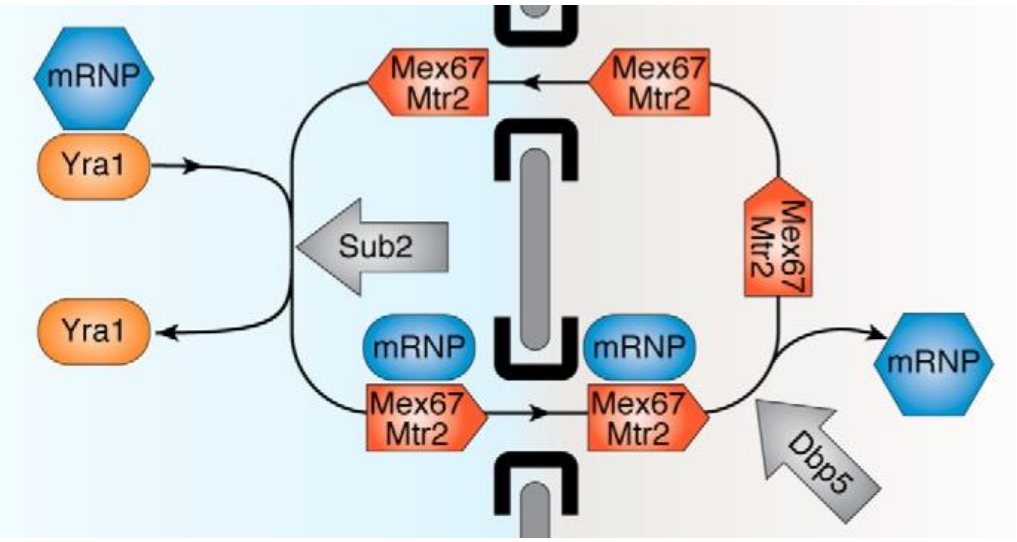


Figure 16: Illustration of the mRNA export factors pathway in yeast. Taken from [110].

1.3. Quality control (QC) of the mRNPs

The mRNP biogenesis machinery is robust, but not infallible. To mitigate eventual errors, cells have evolved a system for mRNP quality control. This system can target and destroy aberrantly formed mRNPs.

1.3.1 Yeast Exosome

The accumulation of aberrant transcripts in the cells is deleterious in two ways. First, the aberrant mRNPs are in competition with correctly formed ones for various factors, both in the nucleus and in the cytoplasm. Second, the translation of an aberrant transcript could lead to the formation of a deleterious protein. Therefore, the degradation of these faulty transcripts is critical for the cell. The degradation of these transcripts is handled by the exosome, a complex that is composed of 12 main proteins, 9 of which are organized in a barrel shaped structure, referred to below as the main core of the exosome. In addition to the barrel, 2 exonucleases are positioned at each extremity of the barrel, Dis3 at the bottom and Rrp6 on top. An additional helicase, Mtr4, is positioned above Rrp6 and unwinds the to-be-degraded transcript. The exosome has multiple purposes, handled by either Rrp6, Dis3 or both [111]. In addition to their overlapping roles, Rrp6 is known to allosterically stimulate Dis3 activity [112].

The main core of the exosome is composed of 9 proteins forming a two-ringed barrel-shaped structure, commonly named Exo9. The upper ring is made of 3 subunits (Rrp4, Rrp40, and Csl4) which are referred to as a “cap”. They possess S1/KH domains similar to those found in RNA-binding proteins. The lower ring is composed of 6 subunits with the fold typical of a bacterial 3′–5′ ribonuclease, RNase PH, but lacking functional active sites. To form the holocomplex that is the exosome (Exo14), the core of the exosome is completed with 5 proteins, 1 situated below the barrel-shape and the others on its upper side. The protein below the core is Rrp44 or Dis3, which is one of the two exonucleases of the complex. The 4 other proteins are Rrp6, the second exonuclease with its two cofactors, Rrp47, needed for Rrp6 stability [113, 114], and Mpp6 [115]. The last protein is Mtr4, the associated helicase of the complex. The association of the latter depends on the presence of both Rrp6 and Rrp47 [116]. Recent studies conducted in-vitro and in-vivo suggest that Mpp6 is another actor able to link Mtr4 to the exosome in humans [117, 118].

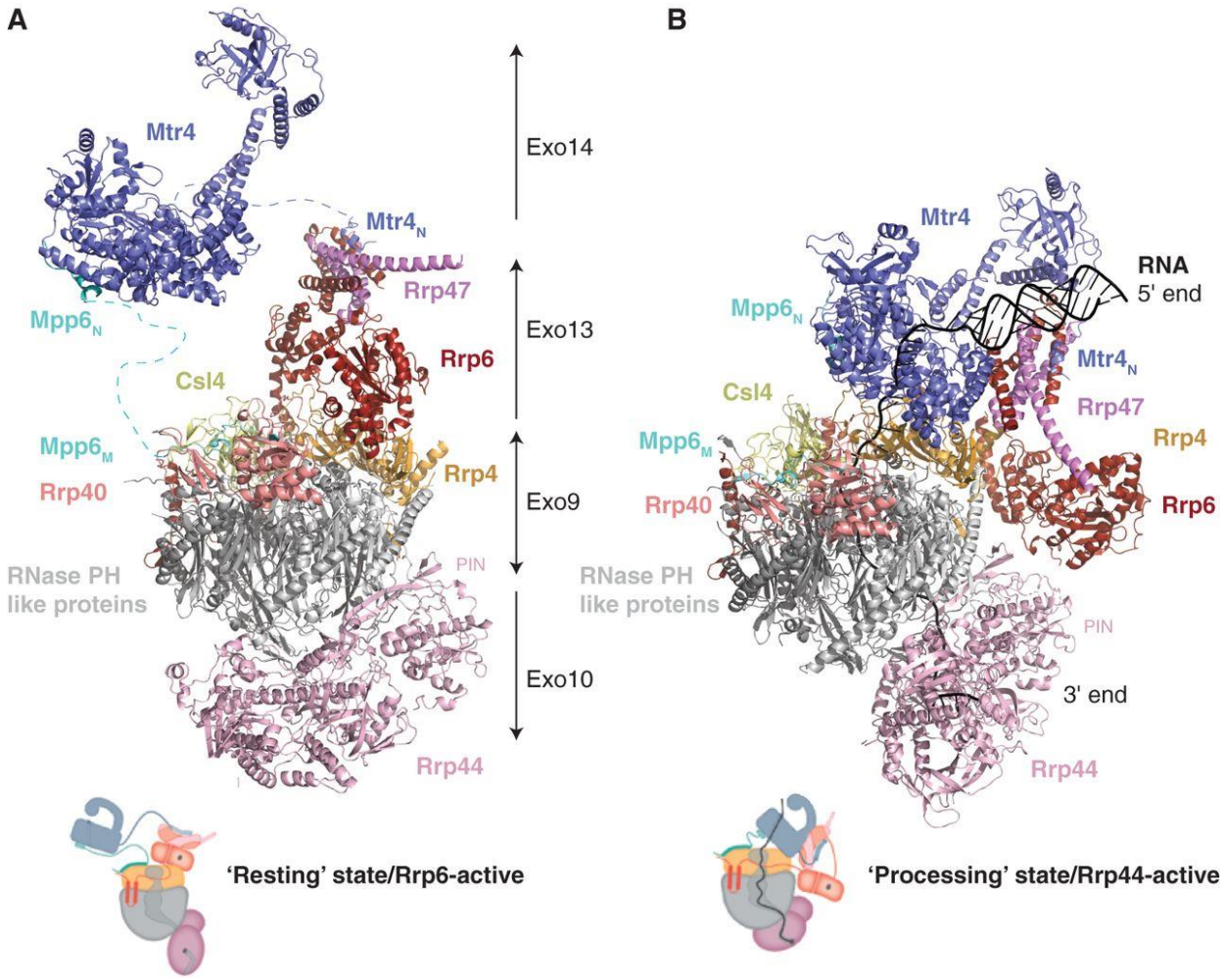


Figure 17: Architecture of the exosome complex. These models are superpositions of several crystal structures. Exo9-10-13-14 are the names used to define the exosome depending on the subunits associated to it. (A) Composite structure of the eukaryotic exosome (cartoon representation) in resting mode. The helicase core of Mtr4 is shown in a hypothetical arrangement. The low-complexity carboxy-terminal sequences of Mpp6 and Rrp6 are flexible and not shown. (B) Composite structure of the eukaryotic exosome in processive mode. Taken from [111]

The Exo9 barrel is traversed by a central channel that will be used for the guidance of the target RNA to the appropriate exonuclease. The helicase activity of Mtr4 is required to unravel the RNA and thread it into the exosome central channel. However, only Dis3 processes the transcripts going through this channel while the transcripts access Rrp6 by passing through the S1/KH cap ring [119]. RNAs can also directly access Dis3 without going through the central channel of the exosome [120]. The importance of this channel for Dis3 and Rrp6 exonuclease activities is debated as some studies support the existence of channel-independent processing pathways

[120-122] while others insist on the importance of this channel for RNA processing [119, 123-125].

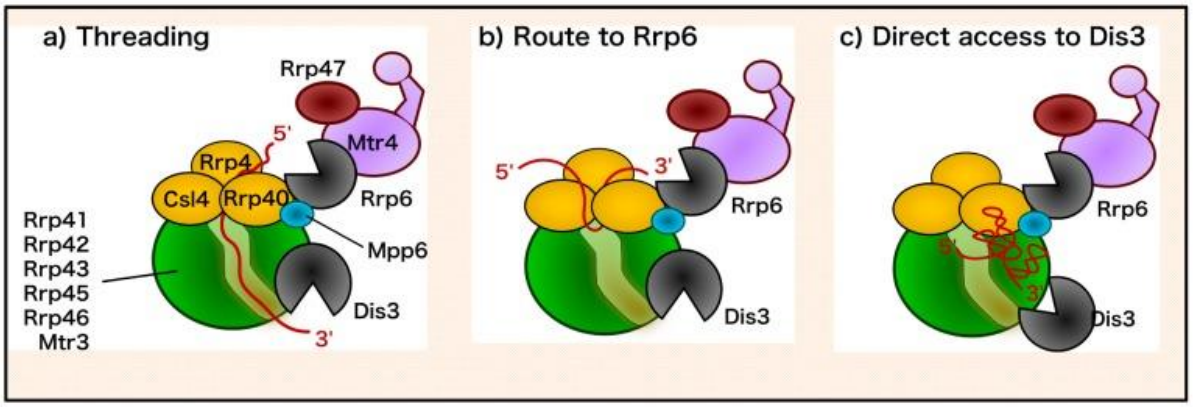


Figure 18: Different representations of the path followed by RNAs in the exosome. (A) Path of the RNA to access Dis3 through the central channel. (B) Access to Rrp6 through the S1/KH cap ring. (C) Access to Dis3 directly. Taken from [126]

As mentioned above, Rrp6 is one of the two 3'-5' exonucleases of the exosome. It is involved in RNA processing and RNA degradation, which includes aberrant RNA elimination. Rrp6 is present in the nucleus in both yeast and human (known as Exosc10 in the latter), although Rrp6 has also been detected in the cytoplasm only in human [127, 128].

Although Rrp6 is not critical for yeast survival, it is required for the 3' end trimming of pre-sn/snoRNA, the degradation of CUT and aberrantly terminated mRNA, the regulation of polyA tail length and the termination of a sub-population of short transcripts [129-131]. Rrp6 is also responsible for the maturation of 5.8S rRNA and, alongside TRAMP5 (Trf5-Air1/2-Mtr4), contributes to the degradation of aberrant pre-rRNAs [132]. By using advanced tiling arrays, some studies showed that Dis3 and Rrp6 have both common and separate roles in the degradation and/or processing of various classes of RNAs [133]. Notably, Rrp6 is linked to the 3' processing of specific ncRNAs, which are not targeted by the canonical polyA-dependent termination mechanism. [134]

Furthermore, some evidences suggest that Rrp6 and Dis3 may retain some effects even when not linked to the exosome core [135-137]. Disruption of Rrp6, Dis3 and Rrp43 showed that each one of them was crucial for the degradation of specific

classes of the RNA population [138]. A recent study associated Rrp6 as a regulator of TRAMP-mediated degradation of tRNA by deadenylation of TRAMP-processed transcripts [125].

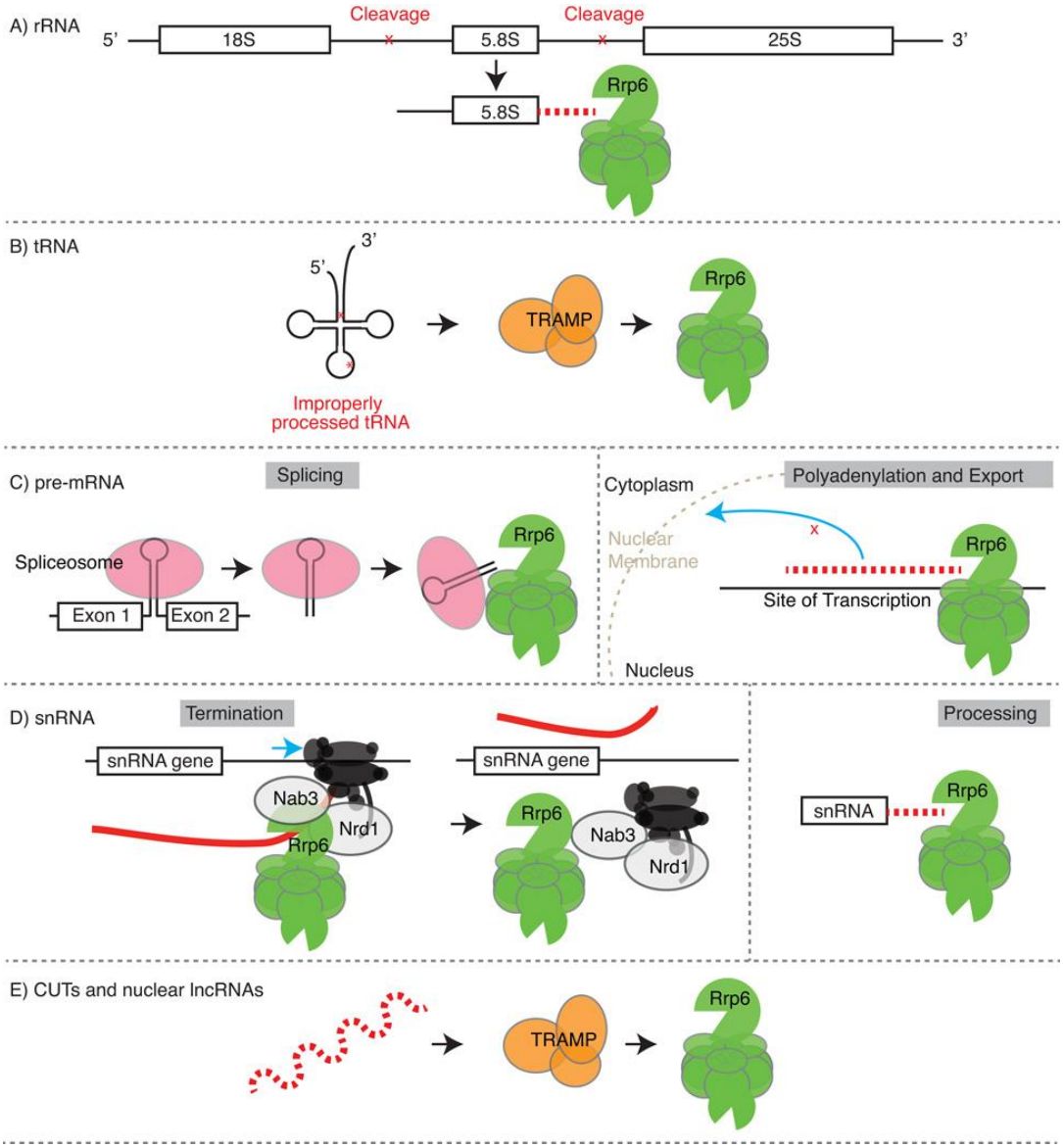


Figure 19: Summary of Rrp6 roles and targets. (A) The 18S, 5.8S, and 25S rRNAs are transcribed as one molecule and cleaved and processed by several nucleases. Specifically, Rrp6 is required for processing of the 3' end of the pre-5.8S product. (B) Improperly processed tRNAs in the nuclease are targeted by the TRAMP complex which adds a short polyA tail and directs the substrate to the exosome for complete degradation by Rrp6. (C, left) Exo11 interacts directly with the spliceosome to degrade introns co-transcriptionally. (C, right) Rrp6 also degrades mRNAs that cannot be exported and accumulate at the transcription site when the accumulation is due to improperly processed 3' ends or inhibition of nuclear export machinery. (D, left) Rrp6 is required for proper termination by Nrd1. (D, right) Rrp6 processing

the 3' end of snRNAs. The extended 3' ends of the pre-snRNA are trimmed back by Rrp6. (E) Cryptic Unstable Transcripts (CUTs) and unstable nuclear lncRNAs are both targeted to the Exo11 by the TRAMP complex. Taken from [139].

Rrp6 structures and domains were intensively studied. Identified domains were summed-up in figure 20 taken from [116]. One interesting feature of Rrp6 is a 100 amino acid lariat, which is able to bind proximal RNA to the exosome channel and enhance RNA decay. [112].

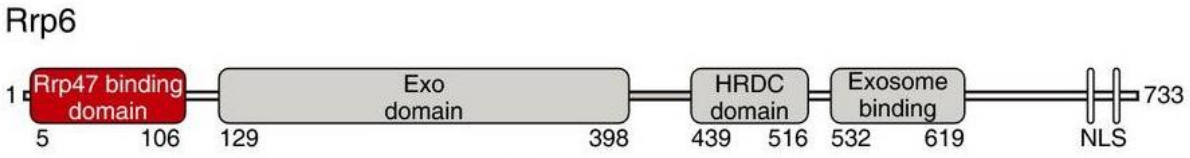


Figure 20: Schematic representation of the domain arrangement of the nuclear exosome cofactor Rrp6 from Saccharomyces cerevisiae. The HRDC (helicase and RNaseD C-terminal) domain is comprised of two orthogonally packed α-hairpin subdomains, and is involved in interactions with DNA and protein. Taken from [116]

Rrp6 interacts with several other proteins, which can be exosome-related or not. As previously said, within the exosome, Rrp6 is in close association with Rrp47. Additionally, Rrp6 also interacts with Mpp6, which is important for 3' end processing of 5.8S rRNA, pre-mRNA/pre-rRNA surveillance and a subpopulation of CUT [140, 141]. Rrp6 also interacts directly with two subunits of the NNS complex, Nab3 and Nrd1 [129, 142]. Interestingly, the interaction between Nrd1 and Rrp6 is linked to the PolII CTD [143]. The role and function of the NNS complex will be detailed below.

Dis3 is the second exonuclease of the exosome. It is a highly conserved 3' to 5' exoribonuclease with additional endonuclease activity [144]. Dis3 is involved in mRNA quality control due to its exonuclease activity [145, 146] and small RNA processing [147, 148]. It hydrolyzes single-stranded RNA in a 3' to 5' direction, releasing one nucleotide at a time and leaving a product a few nucleotides long [149].

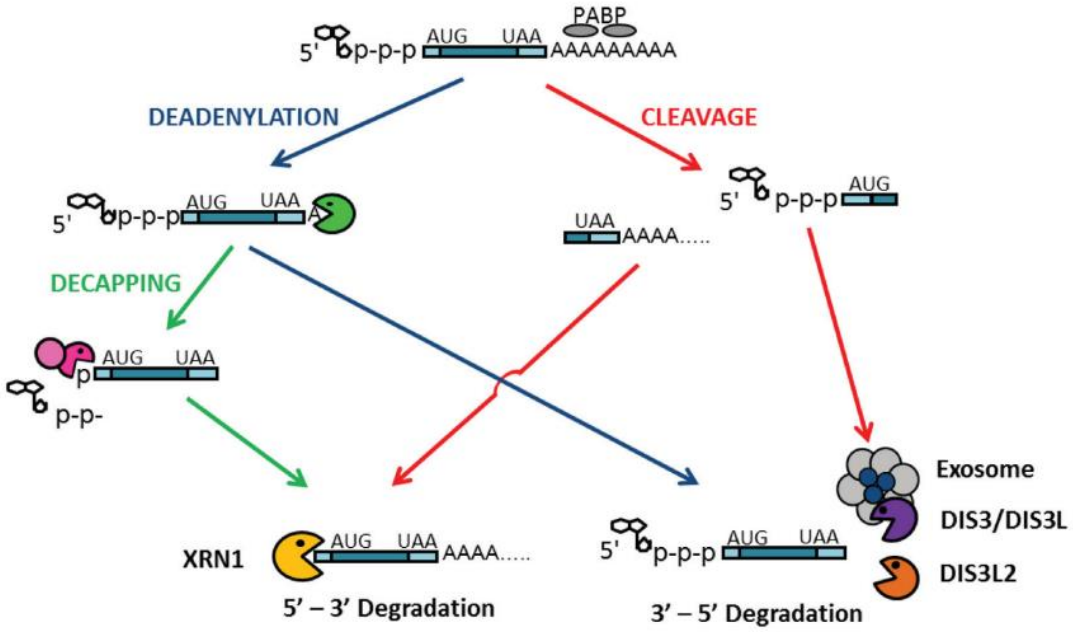


Figure 21: Overview of the mRNA degradation pathways in eukaryotes. Messenger RNAs first undergo removal of the 3' poly-A tail (deadenylation) allowing access for 3' 5' degradation by the exosome complex and DIS3. Following deadenylation the mRNA can undergo removal of the 5' cap (decapping) exposing the mRNA to degradation by the 5' 3' exoribonuclease XRN1. Alternatively mRNAs can undergo endonucleolytic cleavage (e.g., due to RNAi, or nonsense-mediated mRNA decay in some organisms) creating two fragments, each of which is susceptible to either XRN1 or the exosome and the DIS3 paralogues. Taken from [146]

The working model posits that when an RNA or ribonucleoprotein (RNP) substrate is recruited to Mtr4 or to an Mtr4-adaptor complex, the helicase activity of Mtr4 unwinds the RNA substrate, thus progressively extruding the unwound 3′ end toward the top of the exosome's central channel. *In vitro*, the RNA 3′ end will first encounter the active site of Rrp6 and stochastically be degraded in a distributive manner before getting threaded into Exo9 and channeled to the process site of Dis3. *In vivo*, it is suggested that the chemical structure of the 3' end such as 3' phosphate could be a determining factor for the targeting of a transcript by Dis3 or Rrp6 [150].

1.3.2 Targeting co-factors in yeast

If the exosome provides the entire catalytic process of RNA degradation, other players are required to target these transcripts. These players are intricately linked to

the mRNP formation because the same players are responsible for both protection or processing and targeting. For example, Nrd1 is a key protein in the processing of snoRNA and ncRNA but it has also been shown to play a role in the targeting and degradation of aberrant mRNP transcripts within the NNS complex along with the TRAMP complex under stress conditions.

1.3.2.1 TRAMP complex

The first description of TRAMP was made two decades ago in *S.Cerevisiae* and started with the discovery of Trf4, a subunit of TRAMP responsible for tRNA exosome dependent quality control, rapidly followed by a full complex characterization [151-153].TRAMP is a multimeric complex of three different proteins and has several isoforms. It is composed of the helicase Mtr4, a non-canonical poly (A) polymerase Trf4 or Trf5 and a Zn-knuckle RNA binding protein Air1 or Air2 [142, 154]. All these subunits are conserved in humans. Air1 and Air2, which are critical for TRAMP assembly and function, contain five CCHC zinc knuckles that bind to RNA and Trf4 [153, 155]. While *trf5Δ, air1Δ,* and *air2Δ* single mutants are viable, the *trf4Δ* single mutant and *air1Δair2Δ* double mutant are growth impaired, indicating that Trf4 and Air activity is critical for cell function [155, 156]. A recent study showed that there are in fact 3 TRAMP complex variations: Trf4+Air1 (TRAMP4-1), Trf4+Air2 (TRAMP4-2) and Trf5+Air1 (TRAMP5-1). Mtr4 is presumably associated with these complexes. TRAMP5-1 preferentially targets the ITS1 spacer region of 35S pre-rRNA, a characterized exosome substrate for which no AIM domain ribosome synthesis factor has been identified. TRAMP4-2 is more strongly associated with RNAPII transcripts, particularly mRNA 5′ ends, close to the TSS and with CUTs, SUTs, and XUTs. The binding of Air1 and Air2 across mRNA is similar, suggesting that the specificity of the TRAMP complex is mostly dependent on Trf4 and Trf5 [157].

TRAMP assists the nuclear exosome to degrade or trim a large variety of RNA substrates, such as hypomodified initiator tRNA, abnormally processed ribonucleic RNAs (rRNAs), long noncoding RNAs like CUTs (lncRNAs), micro-RNAs (miRNAs) and normal by-products of RNA metabolism such as spliced-out introns [122, 151, 158]. TRAMP is also involved in many other RNA processes such as the maturation steps of precursor RNA and transfer RNA (tRNA) editing [159, 160]. In human, the TRAMP complex is also directly involved in many RNA processing pathways such as

splicing, RNA export and is even indirectly involved in the maintenance of genomic stability [161, 162]. To trigger the recruitment of the exosome to its RNA target, TRAMP adds a polyA tail significantly shorter than the canonical polymerase Pap1 [158]. This short RNA tail is supposed to be an easier substrate for 3'-end degradation.

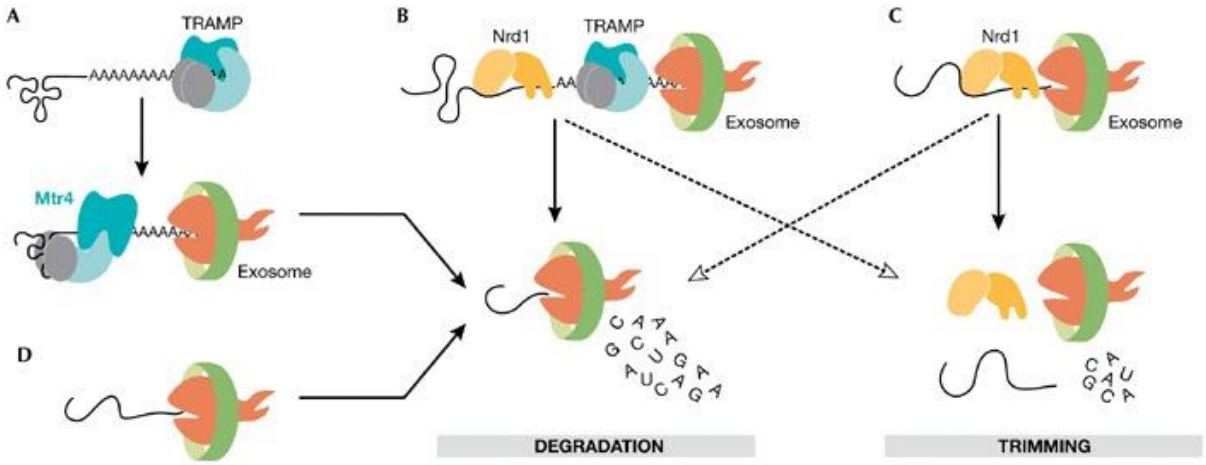


Figure 22: Description of the exosome way of action. (A) The TRAMP complex tags aberrant RNAs with short stretches of oligo(A)s, which initiate RNA digestion by the exosome. (B) Mtr4 helicase of the TRAMP complex unwinds the structured parts of RNAs. The TRAMP complex associates with the Nrd1 complex that binds to short sequence elements on a subset of nuclear RNAs. The interaction between the specific RNA recognition mediated by the Nrd1 complex and the polyadenylation activity mediated by the TRAMP complex acts as the initiation step for RNA degradation by the exosome. (C) The Nrd1 complex can stimulate exosome activity on RNAs with its Nrd1 complex-specific binding sites. This often leads to partial digestion of the RNA (trimming), but it can also cause RNA degradation. (D) The exosome destroys the leftovers of RNA processing, such as the products of endonucleolytic cleavage, apparently by itself. Taken from [163].

In yeast, the mediation of TRAMP on the PolII transcript is coupled to the transcription termination by the NNS complex [134, 164, 165].

1.3.2.2 NNS

The NNS (Nrd1-Nab3-Sen1) complex is a second nuclear exosome cofactor, composed of two short sequence-specific RNA binding proteins, Nrd1 and Nab3, and an RNA/DNA ATP-dependent helicase, Sen1. NNS interacts with RNA PolII and the exosome to stimulate termination and processing/degradation of ncRNAs [129, 166-168]. The NNS subunits bind to different classes of ncRNAs, including snoRNAs,

CUTs, tRNAs, and mRNAs, via either NNS-dependent terminators or via the PolII-CTD [158, 169-171]. NNS-dependent terminators exhibit specific sequence patterns GUA[A/G] and UCUU which can be bound by Nrd1 and Nab3 respectively [172]. All NNS components are essential for the stability of the complex. The depletion of one of the NNS subunits led to a transcriptional read-through of NNS terminators in ncRNAs and elevated levels of CUTs [134, 165]. These unregulated CUTs are called Nrd1-unterminated transcripts (NUTs) [164]. Nrd1 and Nab3 are single RNA recognition motif (RRM)-containing proteins that recognize the RNA sequences GUAG/A and UCUU, respectively. These sequences were thought to be located within the NNS terminators near the 3' end of snoRNA and short PolII transcripts and are important for their termination and degradation [167, 168]. However, a more recent study identified some of these sites with cross-linked NNS-subunits within either the 3' UTR or the 5' end of mRNA transcripts [173]. Our team hypothesized that these sequences could play a role in the targeting of Rho-induced aberrant mRNP by the exosome [25]. Nrd1 and Nab3 also directly interact with one another via Nab3/Nrd1-binding domains [170]. In addition, Nrd1 co-precipitates with the exosome and TRAMP, and more specifically with Rrp6 and Trf4, while Nab3 interacts with Sen1 [80, 129].

In current models for NNS function in PolII-related ncRNA termination, Nrd1 binds to the phosphorylated Ser5 within the C-terminal domain of RNA Pol II during early transcription of an ncRNA and the Nrd1-Nab3 heterodimer recognizes the ncRNA terminator [170, 171, 174]. While the two proteins are on the chromatin, they are able to recruit the Sen1 helicase to dissociate the elongation complex and terminate transcription. Using a mechanism similar to the Rho factor in bacteria, Sen1 dissociates PolII from the chromatin by using its RNA binding and ATP hydrolysis functions. The NNS then leads to the recruitment of TRAMP via an interaction via Nrd1-CTD and Trf4/Trf5. TRAMP then leads the transcript to the exosome for degradation [80]. A recent study showed that only Nrd1 is required for this termination pathway in *S.pombe* [175].

The NNS is also known for its implication in snoRNA termination and processing. Nrd1 and Nab3 interact with NNS sites situated in the 3'UTR of these snoRNAs and trigger the recruitment of the exosome for 3'-end processing [169].

Numerous studies also show that Nrd1 plays a role in the targeting of aberrant mRNPs in yeast [25, 176].

Although Nrd1 and Sen1 homologs exist in humans, none are known for Nab3. Furthermore, different complexes in humans, such as the human TRAMP complex and NEXT, ensure the different roles of NNS in yeast [177].

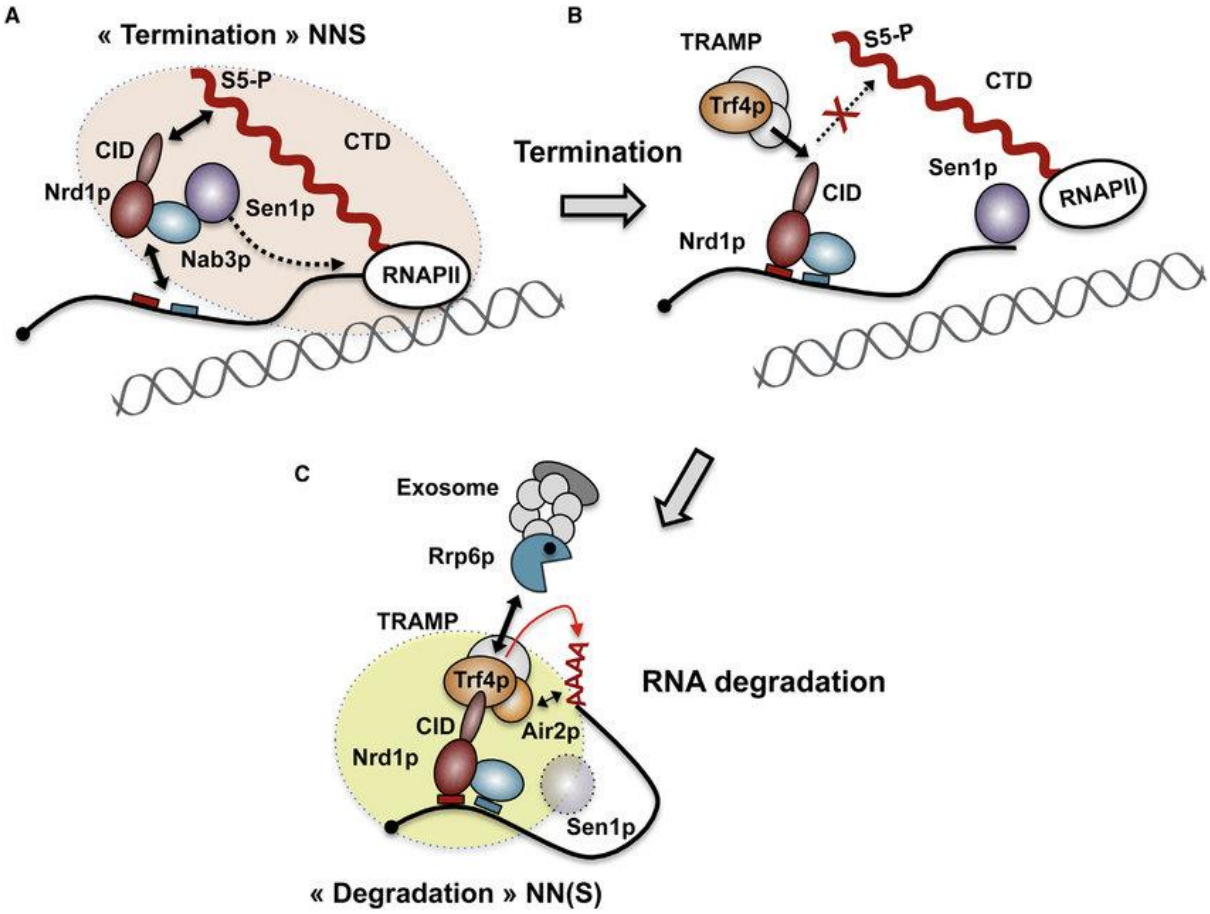


Figure 23: Implications of the NNS in the transcription termination and degradation process. The relationship between PolII CTD and the TRAMP complex is particularly highlighted. Taken from [80].

1.3.3 Human Exosome

Similar to the exosome in yeast, the human exosome harbors two different exonucleases organized around a backbone of nine proteins, six forming a barrel-shaped structure (EXOSC4-9 in human; Rrp41/42/43/45/46/Mtr3 in yeast) and three forming a cap over the barrel (EXOSC1-3 in human; Csl4/Rrp4/Rrp40 in yeast). The

first catalytically active exonuclease named DIS3 in human (Dis3/Rrp44 in yeast) binds the barrel-shaped structure while the second exonuclease EXOSC10/RRP6 binds the barrel cap [178].

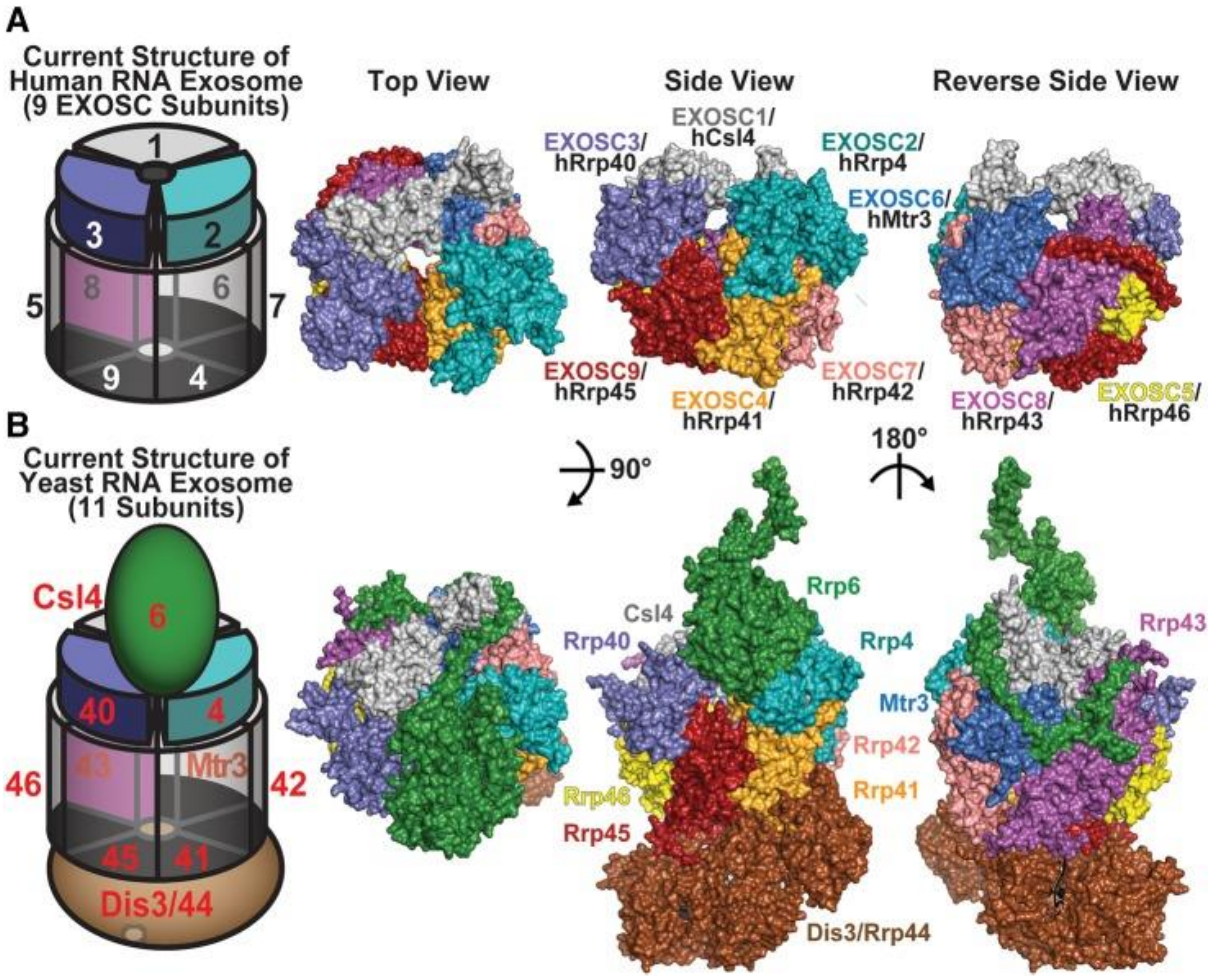


Figure 24: Structure of the RNA exosome, a ribonuclease complex that processes/degrades multiple classes of RNA. (A) Cartoon representation of the nine-subunit human RNA exosome complex that has been solved thus far [179]. (B) Cartoon representation of 11-subunit S. cerevisiae RNA exosome [180] are shown, depicted in top, side, and reverse side views. The color schemes of the human and yeast RNA exosome subunits are identical. Taken from [178]

In contrast to the yeast exosome, which is identical in both nucleus and cytoplasm, two isoforms exist in human, one for the cytoplasm and one for the nucleus. Although nearly identical, the difference between the isoforms resides in the 3'-5' exonuclease DIS3 which exists in 3 homologs, DIS3, DIS3L1 and DIS3L2. Only DIS3 and DIS3L1 are bound to the exosome via a PIN domain which is absent from DIS3L2 thus explaining why the latter is not tightly associated with the exosome [181]. DIS3L1

and DIS3L2 are only located in the cytoplasm while DIS3 is associated with nuclear exosome complexes [182, 183].

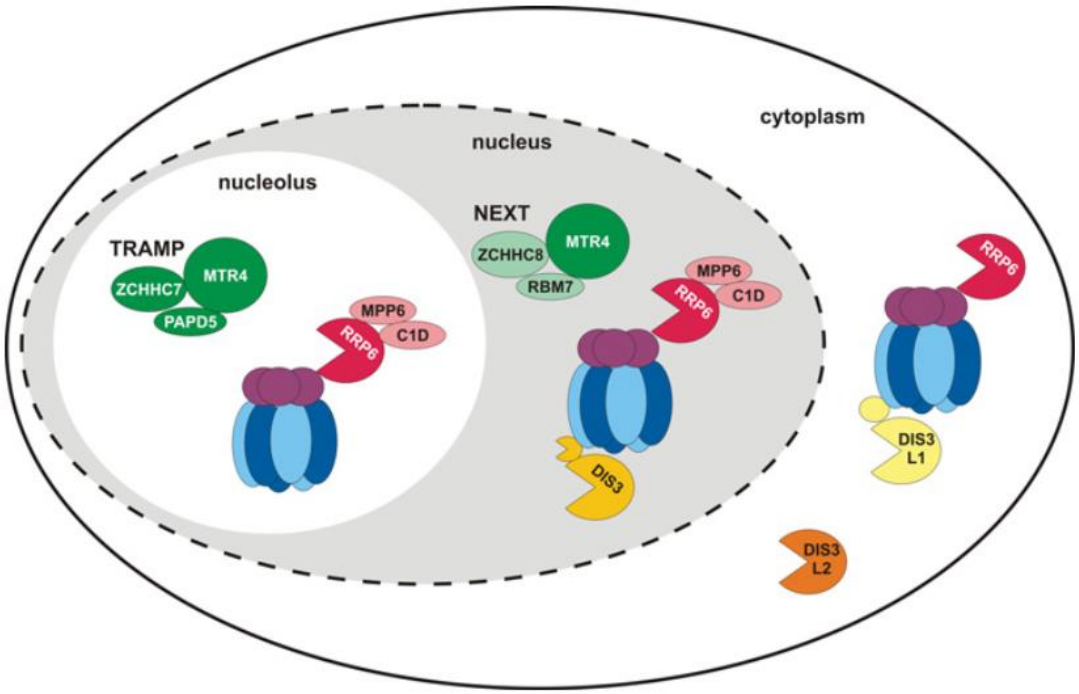


Figure 25: Model of subcellular localization of different forms of the human exosome and cofactors. The core exosome is shown in blue and purple. The PIN domains of DIS3 (active-small 'pacman') and DIS3L1 (inactive-small circle) are shown. Taken from [182]

The roles of the exosome have been studied in yeast for a few decades but more recent studies in higher eukaryotes have added new exosome targets to the already known substrate of the exosome [178]. As developed in the introduction, the RNA exosome produces mature rRNA for the ribosomes but also mature snRNA and snoRNA [147, 184]. Aberrant or unstable ncRNAs such as CUTs in yeast and promoter upstream transcript (PROMPT) in humans are also targeted [156, 185]. In addition, the exosome also targets upstream antisense RNA (uaRNAs) and transcription start site-associated RNA in mice [186, 187]. Finally, the RNA exosome targets aberrant mRNA but also prematurely terminated transcripts from coding genes in humans [178, 188].

The nuclear exosome includes cofactors conserved between yeast and humans. In yeast, Rrp6 requires Rrp47 to be stable [113, 114]. Both proteins recruit Mpp6, which then recruits the helicase Mtr4 [117, 118]. Interestingly, each of these

proteins has an orthologue in the human cell: C1D (Rrp47), MPP6 (Mpp6) and MTREX or hMTR4 (Mtr4) [140, 189, 190].

In addition to these intrinsic cofactors, the exosome collaborates with three different complexes in humans : an homolog of the TRAMP complex (TRAMP-like complex), the poly(A) tail exosome targeting complex (PAXT) and the nuclear exosome targeting complex (NEXT) [162]. These co-factors are each located in a different cellular compartment and therefore determine which transcript will be targeted by the exosome.

1.3.4 Exosome cofactors in human

In yeast, the TRAMP complex is a cofactor guiding the exosome to a subset of RNA targets. In human, the Air1/Air2 orthologue is the zinc-knuckles protein ZCCHC7 and the Trf4/Trf5 orthologue is the non-canonical polyA polymerase PAPD5. Together with MTREX, these proteins form the human TRAMP-like complex [162]. In contrast to its yeast counterpart, this TRAMP-like complex is considered only effective in the nucleoli, at least under normal conditions. This is due to the strictly nucleolar function of ZCCHC7 [191]. Similarly to the yeast TRAMP complex, its human homolog is responsible for the poly-adenylation of snoRNA and aberrant pre-rRNA [191, 192]. In addition, TRAMP-like complexes associate with various splicing factors such as snRNP (small nuclear ribonucleoproteins) although the impact of these interactions needs further exploration [193].

Outside of the nucleoli, two other cofactors associate with the exosome: the poly(A) tail exosome targeting complex (PAXT) and the nuclear exosome targeting complex (NEXT). The first one, PAXT, is a complex of two proteins, MTREX and ZFC3H1 [126, 194]. Additionally, PAXT associates more transiently with the poly(A) binding protein PABPN1, the Zn-finger protein ZC3H3, and presumably one of the two RNA binding protein paralogs RBM26 and RBM27 [194, 195]. The second one, NEXT, is composed of the helicase MTREX, the Zn-knuckle protein ZCCHC8 and the RNA binding protein RBM7 [191]. NEXT targets promoter upstream transcripts (PROMPTs), enhancer RNAs (eRNAs), 3′ extended snRNA, and intronic RNA while PAXT targets a large variety of long non coding RNA from spliced transcripts from snoRNA host genes and prematurely terminated transcripts deriving from intronic poly-adenylation sites

[196]. Interestingly, both possess high binding affinities with the cap-binding complex, suggesting a preference for capped transcripts [194, 197]. However, due to their preferred targets, it is admitted that NEXT targets rather short, unprocessed and less abundant transcripts. In addition, one common feature of these transcripts is that they do not harbor a polyA tail (pA- transcripts). On the other hand, PAXT targets longer transcripts with polyA tails (pA+ transcripts) [194].

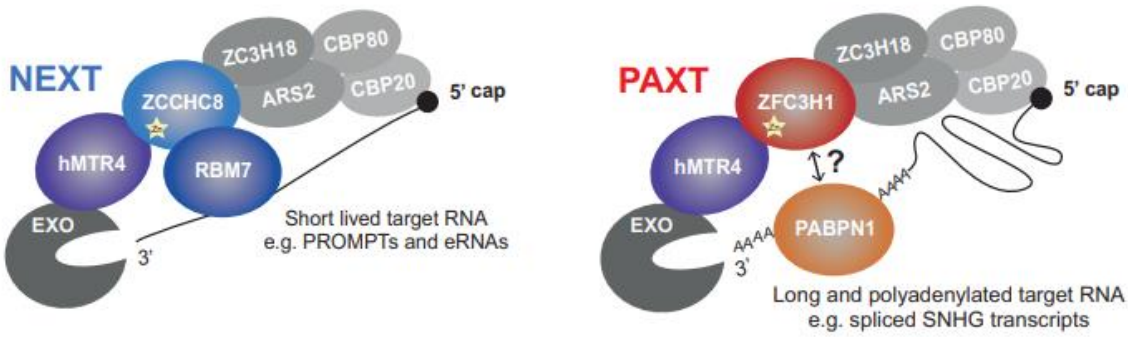


Figure 26: Schematic comparison of protein-protein links within the NEXT complex (left) and the PAXT connection (right). While both NEXT and PAXT pathways appear capable of detecting capped RNA by virtue of their physical linkages to the CBC, the different RNA-binding proteins (RBM7 for NEXT and PABPN1 for PAXT) discriminate their specificities. A question mark indicates that the ZFC3H1-PABPN1 linkage might not be direct. Taken from [194].

However, recent genome-wide analysis of NEXT and PAXT targets suggested that a high portion of the RNA population targeted by NEXT could be handled to the exosome by PAXT, especially as a backup if NEXT is non-functional. Notably, the transfer of a target RNA from the NEXT complex to the PAXT complex requires the polyadenylation of said RNA. Moreover, a large number of exosome targeted loci produce both NEXT and PAXT sensitive RNA isoforms, showing that both pathways are deeply intertwined [196].

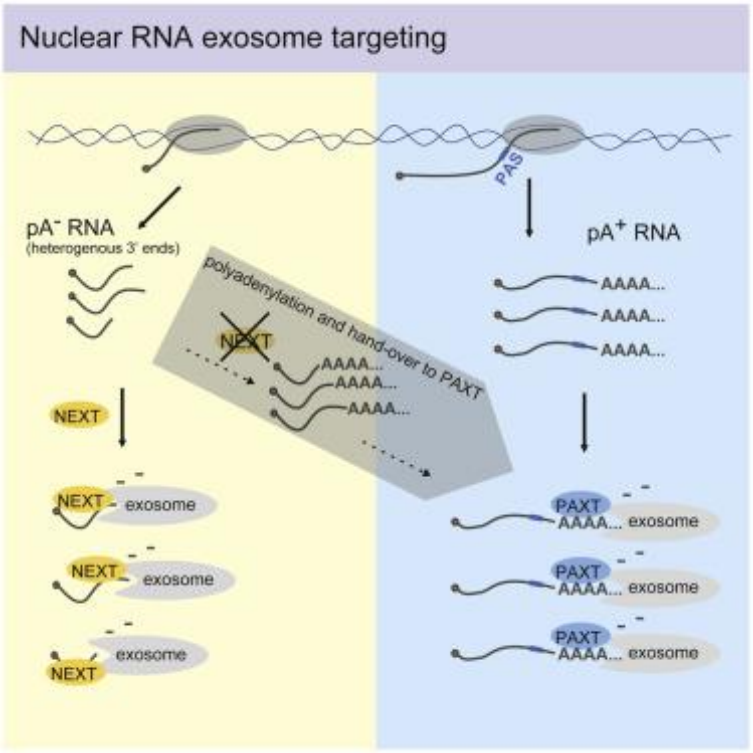


Figure 27: The nuclear exosome adaptor NEXT targets poly(A)- RNAs with poorly defined 3' ends, whereas the PAXT connection targets poly(A)+ RNAs derived from canonical poly(A) sites. NEXT substrates become polyadenylated in the absence of NEXT, causing fail-safe RNA decay via PAXT. Graphic abstract from [196].

The involvement of these cofactors and the exosome in the degradation of Rho-induced aberrant transcripts could provide many insights about RNA degradation regulation in humans.

1.4 Experimental approach

The study of QC mechanisms is arduous because the number of aberrant mRNPs generated under physiological conditions is very small. To increase this number, we used a factor named Rho to induce the formation of aberrant mRNP without influencing the QC system by deleting one or more of its components as it has been done in previous work by knocking down the Tho-Sub2 complex. We have coupled Rho induction with ChIP-seq, allowing us to record events occurring on a genome-wide scale. Since I have devoted a large part of my PhD to learning NGS data analysis, I will discuss both Rho and the emerging field of NGS and its applications.

1.4.1 Rho

Discovered in 1969, Rho is a homohexameric complex commonly found in bacteria [198]. In prokaryotes, it plays a role in transcription termination, being responsible for around 20%-30% of the transcription termination event in bacteria [199] and about 50% in E.coli [200]. Aside from this main role, Rho also suppresses antisense transcription [201], regulates the transcript population when translation is impaired [202]. Finally, it resolves conflicts between transcription and replication machineries [203].

Rho interacts with C-rich, unstructured regions of the RNA called *rut* (for Rho utilization) [204]. RNA interacts with a first series of Rho binding sites on the surface of the hexamer. This interaction allows Rho to open briefly, allowing the 3'-end of the RNA to pass through the center [205]. Its fixation on the transcripts triggers the activation of its ATP-ase dependent translocase activity, which leads Rho to pass along the transcript from 5' to 3' [206]. In addition to its translocase ability, Rho possesses a DNA/RNA helicase ability crucial to the termination of some transcripts. Though it is commonly admitted that Rho is recruited onto the nascent transcript via a *rut* site, the way it reaches the transcription bubble is blurrier. The most accepted model suggests that Rho exerts its translocase activity while still being bound to the rut site, pulling the RNA in a zipper-like fashion [200, 207]. The accession of the *rut* sites is modulated by a variety of signals like ions, metabolites, amino-acid availability, small RNA and proteins [208]. They can modulate Rho-termination by facilitating *rut* accessibility by unfolding RNA secondary structures [209]. Since transcription and translation are coupled in bacteria, these modulators can also trigger translation termination, allowing Rho to access the transcript and induce transcription termination [210].

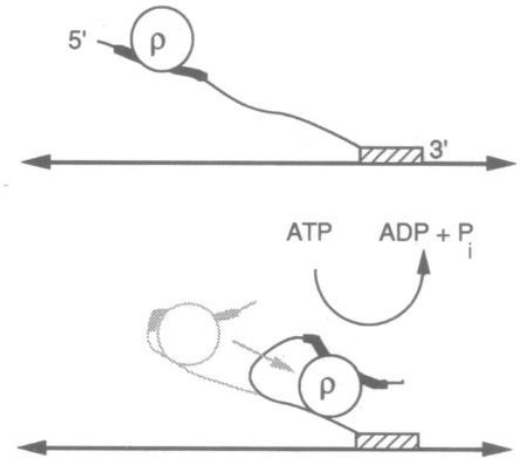


Figure 28: Rho tethering along the nascent transcript does not lead to its dissociation from the rut site. Taken from [207]

The way Rho could trigger the stop of transcription in the kinetic coupling model is also debated [200]. The constraints imposed by its helicase activity on the RNA/DNA hybrid within the transcription bubble could promote the termination of the transcription [211]. On the other hand, the brute force imposed by the translocase activity of Rho could be sufficient to disrupt the DNA/RNA hybrid within the transcription bubble, pulling the nascent RNA from the RNA polymerase and the transcription bubble [212]. This hypothesis is further strengthened by the fact that Rho is able to dissociate streptavidin from biotinylated RNA [204]. In addition to its ability to disengage the polymerase from its template, Rho is known to be able to compete with other proteins for RNA binding, a characteristic used as a transcription regulation mechanism in prokaryotes [213].

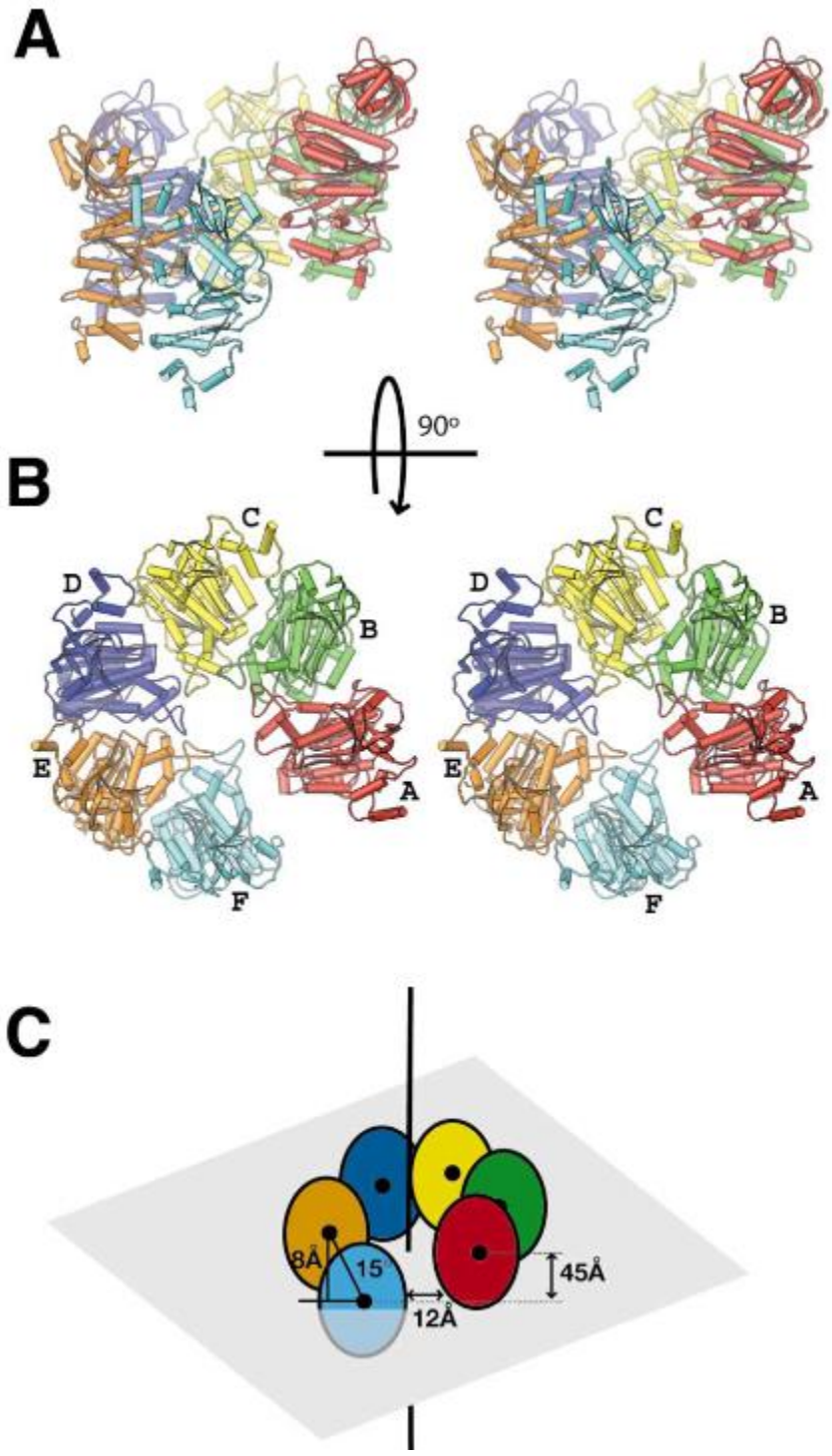


Figure 29: View of the Rho hexamer from the front (A) and side (B). ABCDEF each represents one identical subunit of Rho. Schematic of the relative rise and offset of adjacent Rho subunits as they wind about the pseudo-6-fold axis of the ring (vertical line). Ovals are colored according to the color scheme in (A) and (B). The gap between monomers A and F is 12Å, and the helical pitch is 45Å as represented in (C). Taken from [214].

As mentioned above, Rho is able to disrupt strong RNA-protein interactions. Since a large number of proteins are implicated in mRNP formation in eukaryotes, Rachid Rahmouni's team sought to use this property of Rho as an inductor of aberrant mRNA biogenesis. Subsequently, a model was developed in which Rho was induced in *S.cerevisiae* under a TetO7, doxycycline-repressed promoter [25, 113, 176, 215-217]. The same model was used to create a direct application of the aberrant-mRNP-biogenesis inducing effect of Rho, in which it is used as a screening tool for the discovery of new bacteria-targeting drugs [218].

The last study from the team demonstrated that Rho has a wide variety of targets. Using NGS techniques, this study showed that roughly one transcript out of 5 were affected by Rho in yeast. A further effect of Rho induction in yeast is a growth defect that is associated with the degradation of Rho-induced aberrant mRNP. This factor was used to identify proteins implicated in the targeting and degradation of aberrant mRNP [113, 176, 216]. Recently, we showed that upon mRNP biogenesis perturbation, some factors responsible for ncRNA processing in normal conditions were relocated to mRNA coding loci [25].

Previous work from the team used the same system to highlight the actors required for Rho-induced aberrant mRNP removal, such as Mpp6 or TRAMP subunit Air2 [113]. Using this tool, the team also highlighted a possible parallel degradation pathway of Rho-induced aberrant mRNP mediated by the exonuclease 5'-3' Rat1, implicated in the transcription termination models currently proposed by the scientific community [216].

Finally, during my PhD, I showed that in parallel of these events, mRNP associated protein such as Tho2, from the THO complex, can have a side effect if the transcript they bind to is affected by Rho. I showed that they are able to recruit degradation factors such as Rrp6 onto the affected transcript, leading to its degradation.

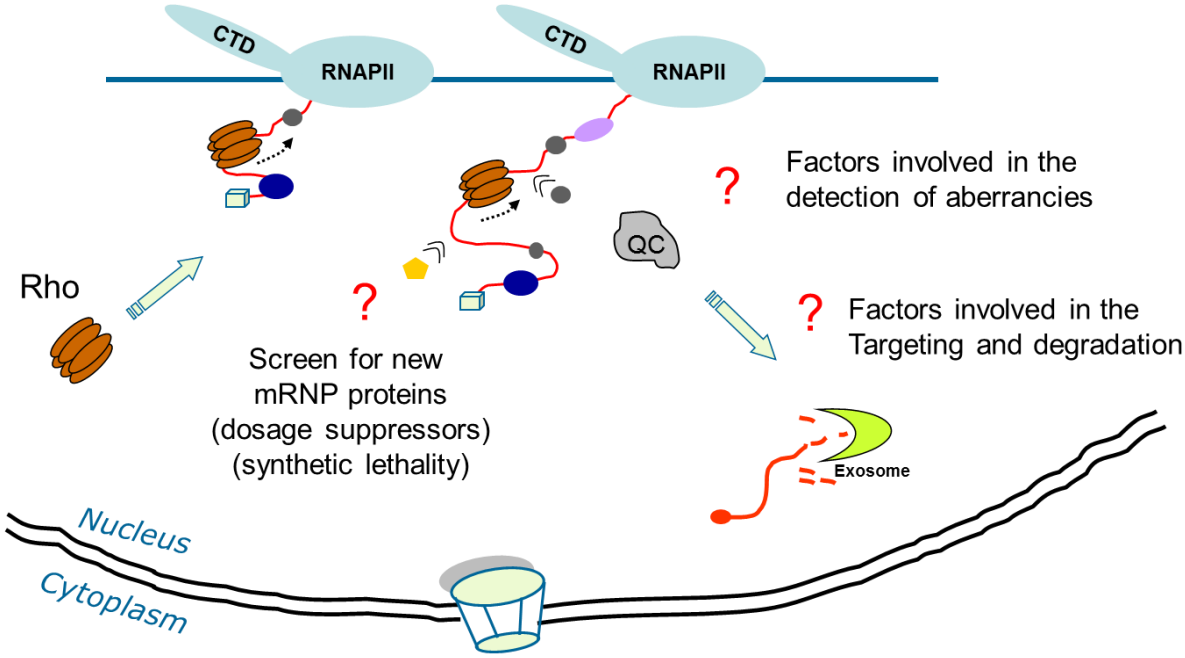


Figure 30: A model of the function of Rho.

1.4.2 Bioinformatics

1.4.2.1 Whole-genome approach with high-throughput sequencing (HTS)

In recent years, the analysis in the genomic and transcriptomic fields has been revolutionized by the rise of high throughput sequencing techniques (HTS), which gave the opportunity to reach genome-wide-scale analysis, allowing the investigation of vast populations of transcripts. This HTS application is usable to study nucleic acid-protein interaction, expression, identification of single nucleotide variants (SNP), RNA structures and epigenetics [219-222].

The HTS principle works with extracts of nucleic acids and their sequencing. The targets of the extractions and eventual transcript treatment can allow different analyses. For example, total mRNA extraction is useful to measure the expression of each transcript. On the other hand, the sequencing of immunoprecipitated chromatin allows the pinpointing of the sites of protein-chromatin interaction of a protein of interest [223]. The modification of transcripts before retro-transcription with SHAPE reagents (Selective 2′-Hydroxyl Acylation analyzed by Primer Extension) allows

bioinformaticians to predict RNA structure with more accuracy than simple mathematical models [222].

After extraction, the transcripts are retro-transcribed and sequenced. Figure 33 describes the sequencing technology used by Illumina.

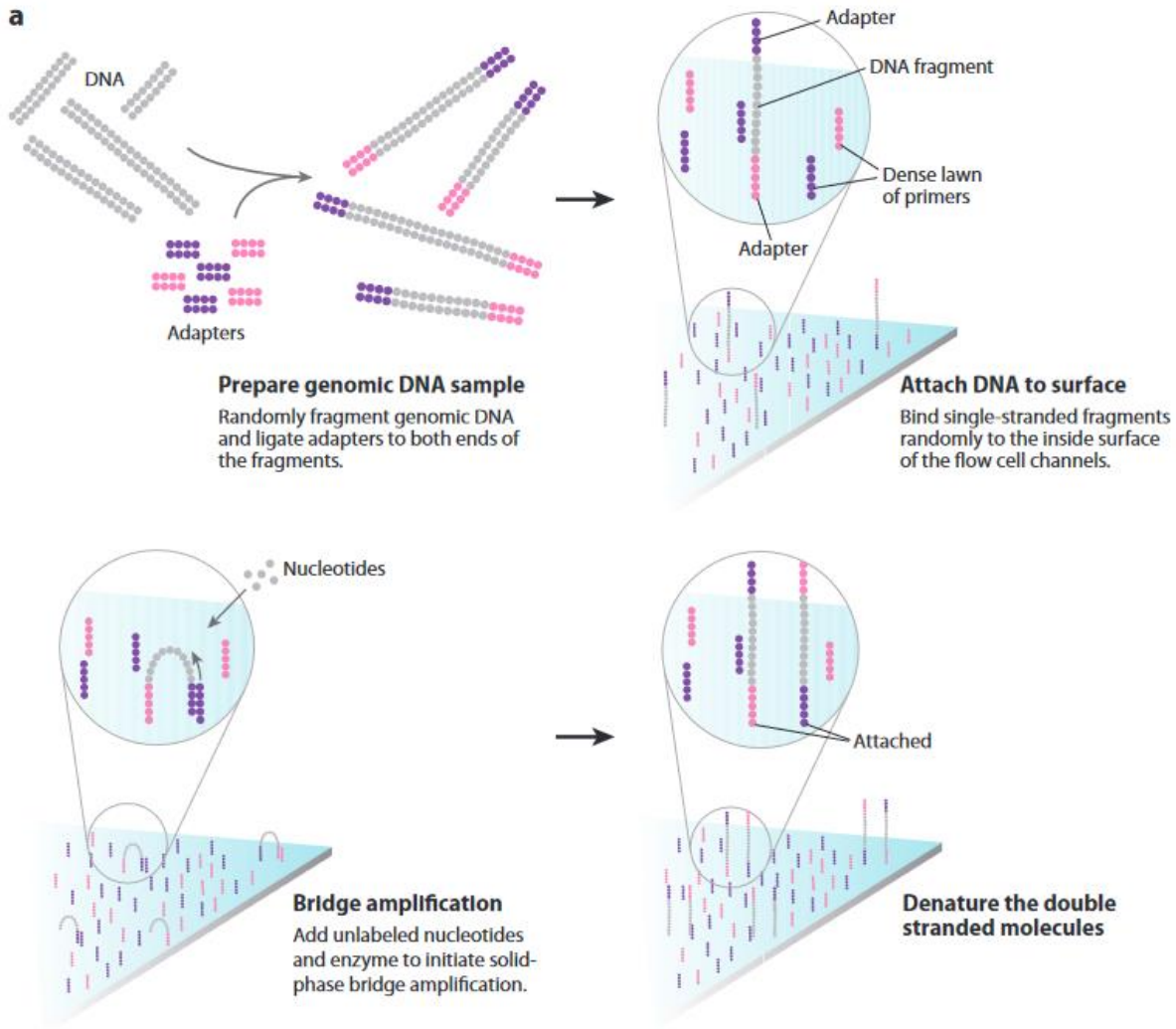


Figure 31: The Illumina sequencing-by-synthesis approach. Cluster strands created by bridge amplification are primed and all four fluorescently labeled, 3'-OH blocked nucleotides are added to the flow cell with DNA polymerase. The cluster strand is extended by one nucleotide. Following the incorporation step, the unused nucleotides and DNA polymerase molecules are washed away, a scan buffer is added to the flow cell, and the optics system scans each lane of the flow cell by imaging units called tiles. Once imaging is completed, chemicals that effect the cleavage of the fluorescent labels and the 3'-OH blocking groups are added to the flow cell, which prepares the cluster strands for another round of fluorescent nucleotide incorporation. Taken from [224].

The obtained set of millions of sequences is then mapped onto a reference genome, which is the entire sequence of the genome of an organism of interest. This reference genome is annotated with all known genomic features (coding and non-coding loci for example). By counting the number of mapped sequences per feature, we obtain a count table, which will serve to quantify the expression level of the different features (if we have a total extract). For ChIPed extracts, we can perform a ratio with/without immunoprecipitation for each base or for a base group, in a similar manner to the traditional ChIP analysis. The results of this analysis show a relatively enriched region corresponding to the site of interaction between chromatin and protein. Further post-analysis possibilities are motif discovery algorithms, sequence comparisons, eventually GO-term.

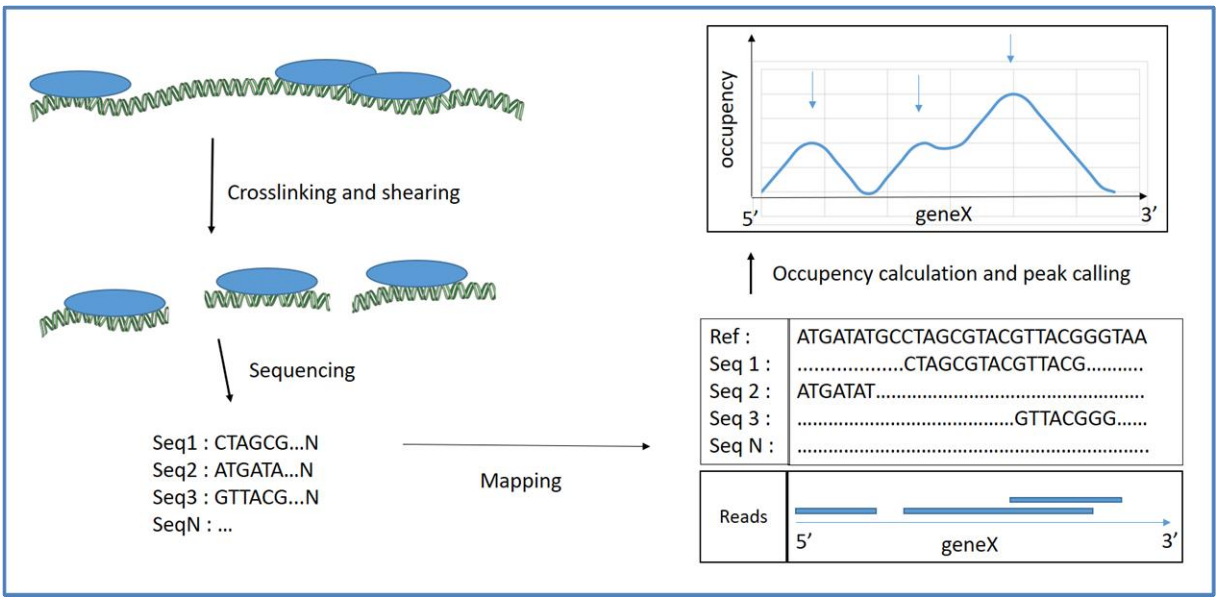


Figure 32: ChIP-seq simplified steps. After crosslinking protein and RNA, the fixed protein-RNA molecule is sheared into fragments of about 200 to 300 nts. These fragments are sequenced following the Illumina protocol shown in the previous figure. The resulting sequence are then assigned based on the features of a reference genome and then counted. The ratio between ChIPed (treated) and input counts gives us two pieces of information, namely occupancy and peak localizations.

1.4.2.2 Statistical models and Deep Learning

Despite the revolution HTS brought to the genomic field, it has its limitations, which depend on the intended application of HTS. For example, the recurrent challenges faced with *de novo* genome assembly, a typical application in HTS are long

repeats, heterozygosity, data accuracy, and measuring assembly quality [225]. The construction of a functional model may also require additional statistical analysis and/or signal optimization and normalization. To address this objective, researchers incorporated statistical models into their analyses. In addition, other models and statistical methods like analysis of variance (ANOVA), linear regression or generalized additive models can be used to analyze HTS data. Other statistical tools include false discovery ratio (FDR) control and Familly-Wise Error rate control, which allow the scientist to appreciate the quality of the data generated [226]. The use of more complex models to complement this basic test is also possible. For example, the Hidden Markov model (HMM) is used on continuous data, like a genome' sequence, to optimize the discovery of an event on a base level. Other examples of known applications of HMM are gene finding, multiple sequence alignment and regulatory site identification [227].

In addition to these statistical models, an IA-based approach called Machine Learning (ML) also recently emerged. The idea is to make a program learn from the data, especially by extracting variables following a mathematical model from the training data and using these variables to find characteristics in "real" data. To be more precise, the data we generate in genomics is processed following a sub-class of ML called Deep Learning. It can be summarized, and oversimplified as a global application of neural networks (another subtype of ML) to determine interesting properties or variables from a dataset [228]. Although I did not use deep learning techniques during my PhD, the development of this field made deep learning an essential component of a bioinformatician's toolbox. The applications are multiple and some are summarized in the figure below.

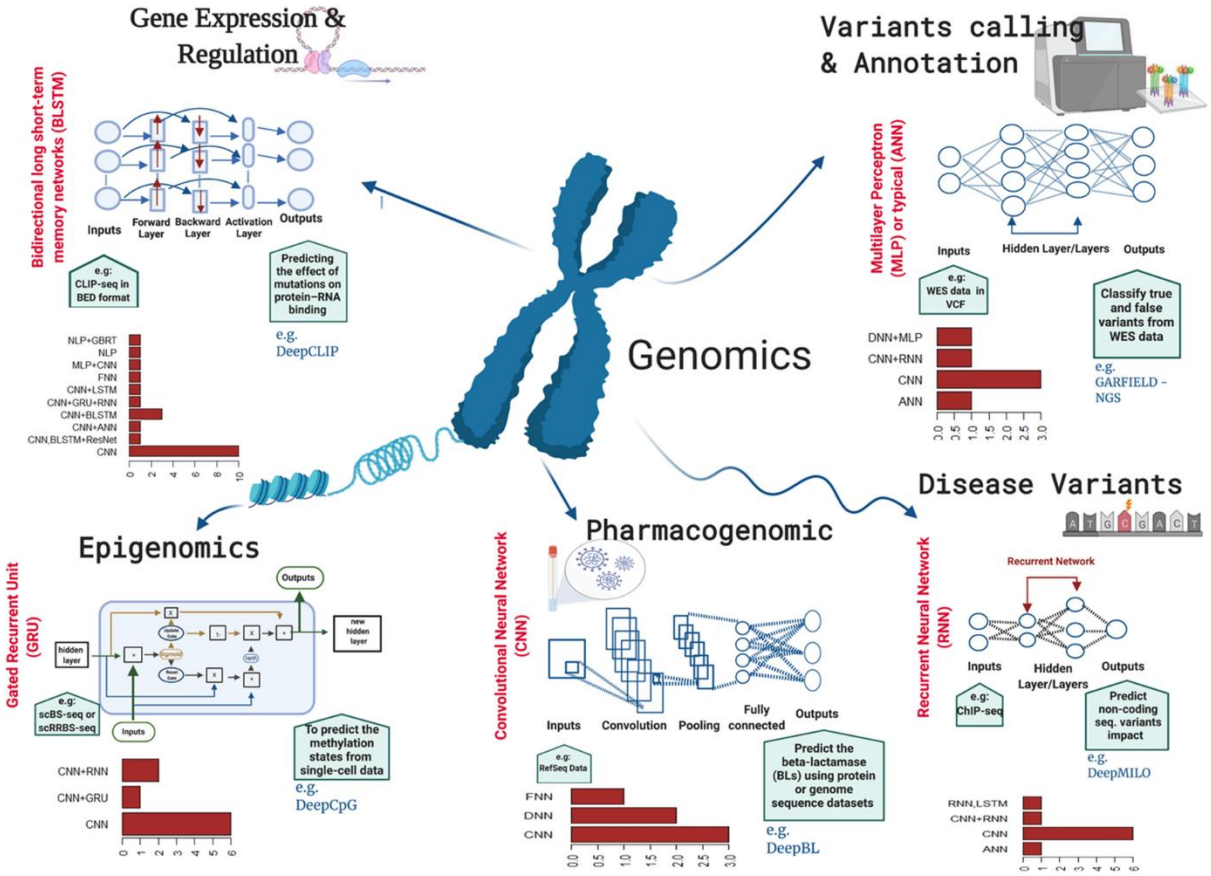


Figure 33: Taken from [228]. Deep learning applications in genomics. This figure represents the application of deep learning tools in five major subareas of genomics. One example of the deep learning tool and underlying network architecture is shown for each of the genomic subareas, and its input data type and the predictive output are briefly mentioned. Each bar plot depicts the frequency of the most commonly used deep learning algorithms underlying deep learning tools in that subarea of genomics.

In my problem, we wanted to evaluate the behavior of the subunits of the THO complex in the case of mRNP biogenesis perturbation. To ensure that we have a global view of the mechanism, we performed a series of Chromatin ImmunoPrecipitation followed by deep sequencing (ChIP-seq). This technique allows us to pinpoint the location of proteins of interest and quantify their interaction with chromatin.

1.5 Objectives

The functioning of the degradation machinery is well documented. However, the detection and targeting of aberrant transcripts is still a shady area. The difficulty of studying this targeting mechanism is due to two factors. First, the QC of mRNPs encompasses a variety of protein complexes that can interact with each other directly

or indirectly through the RNA transcript or via the PolII. These interactions mostly depend on the transcribed gene, as previously identified by our team [25]. Thus, a study focusing on a single example gene might compromise our understanding of the entire targeting mechanism. Second, the PolII and the subsequent mRNP assembly are resilient systems that are quite unlikely to produce aberrant mRNPs. This leaves a very narrow window of opportunity to study the degradation of aberrant mRNPs, something that we are currently technologically unable to address, even with NGS. Therefore, the use of a model that leads to the generation of aberrant transcripts, is almost inevitable. Similar observations could be obtained with the deletions of processing and/or export factors but the possibility to overlook interesting results due to these deletions is to be considered.

To answer these two factors, we chose two approaches. The first is the use of NGS technologies to avoid single locus bias. Using whole genome ChIP-seq, we can examine the interaction of our proteins of interest with all nascent RNAs that are still attached to chromatin by the transcription machinery. This approach would need controls to ensure that nascent RNAs and not only chromatin are targeted. One example of such control would be to evaluate signal loss upon RNAse treatment. NGS results are also more reliable when it comes to building a model as they capture biological signals globally. The second approach is to use the bacterial Rho factor to induce the formation of aberrant mRNP in eukaryotic cells. Rho allows us to extend the time window in which we can study aberrant mRNP targeting. Used together, these approaches are a perfect tool to determine the genome-wide behavior of QC-involved complexes.

My topic has its origins in previous works by my team. Having used Rho for more than a decade, we know that Rho induces a variety of physiological phenomena in yeast, such as slower growth and a high recruitment of Rrp6 on actively transcribed loci throughout the genome. In the literature, the same effects were observed in mutants for the THO/Sub2 complex. We hypothesized that Rho might have an effect on THO. Thus, we studied the effect of Rho on THO recruitment on aberrant transcripts, to see if THO or any of its subunits played a role in targeting aberrant mRNPs.

During my PhD, I tried to answer two other questions. The first was the *rrp6Δ* phenotypic growth defect and heat sensitivity. In collaboration with Dr. Igor Stuparevic's team in Zagreb, we showed that Rrp6 regulates the expression of a number of genes coding for crucial cell wall proteins. The goal of our team in this project was to confirm the effects of Rrp6 deletion by metagenomics analysis of the expression of cell wall related genes depending on their ontology. We also monitored the effects of this deletion on the expression of antisense or proximal ncRNA.

The second task was to analyze the NGS data generated from our transposition of the Rho system from yeast to human cells. Consistent with previous work by my team, my aim was first to identify a subpopulation of transcripts affected by Rho, proving that Rho would be a suitable model to study aberrant transcript degradation in humans. Then, using Rho, I tried to single out transcripts potentially retained in nuclear bodies.

The next part of this thesis manuscript will describe how I answered these hypotheses. I will first present the two completed publications and I will finish with a third part about the current, ongoing work.

2. Results

2.1 Tho2 moonlights in the yeast co-transcriptional mRNP quality control by targeting aberrant mRNPs to Rrp6

2.1.1 Introduction

The importance of THO in the processing and export of mRNP has been studied for more than two decades [106, 229-231]. Its role is so important that the deletion of the THO-Sub2 complex has been used as a model to study the quality control system of mRNP in yeast [105]. We were initially interested in the similarities between the THO-Sub2 deletion model and our Rho model. We suspected that Rho might affect the function of the THO complex. Therefore, we monitored the recruitment of THO and each of its subunits on the chromatin upon Rho induction in *S.cerevisiae.* For this purpose, I used the Rho model in conjunction with bioinformatics, similar to [25]. I showed that Rho indeed disrupted the usual recruitment of the THO complex to

chromatin. I also found that the Tho2 subunit had a completely different recruitment pattern upon Rho induction. The recruitment of Tho2 to chromatin was also strongly correlated with the recruitment of Rrp6 at the same genomic loci. We could establish a functional link between Tho2 and Rrp6 but this interaction could be direct or indirect and remains to be explored.

2.1.2 Manuscript

Tho2 is critical for the recruitment of Rrp6 to chromatin in response to perturbed mRNP biogenesis

Abbreviated title: Tho2 moonlights in yeast mRNP quality control

Valentin Beauvais[1], Kévin Moreau[1], Bojan Žunar[2], Nadège Hervouet-Coste[1], Ana Novačić[2], Aurélia Le Dantec[1], Michael Primig[3], Christine Mosrin-Huaman[1], Igor Stuparević[2,*] and A. Rachid Rahmouni[1,†]

[1]Centre de Biophysique Moléculaire, UPR 4301 du CNRS, 45071 Orléans, France

[2]Laboratory of Biochemistry, Department of Chemistry and Biochemistry, Faculty of Food Technology and Biotechnology, University of Zagreb, Zagreb, Croatia

[3]Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMR_S 1085, F-2 Rennes, France

[†]Deceased April 2021

*Corresponding author: istuparevic@pbf.hr

Abstract

The eukaryotic THO complex coordinates the assembly of so-called messenger RNA-ribonucleoprotein particles (mRNPs), a process that involves co-transcriptional coating of nascent mRNAs with proteins. Once formed, mRNPs undergo a quality control step that marks them either for active transport to the cytoplasm, or Rrp6/RNA exosome-mediated degradation in the nucleus. However, the mechanism behind the quality control of nascent mRNPs is still unclear. We investigated the co-transcriptional

quality control of mRNPs in budding yeast by expressing the bacterial Rho helicase, which globally perturbs yeast mRNP formation. We examined the genome-wide binding profiles of the THO complex subunits Tho2, Thp2, Hpr1, and Mft1 upon perturbation of the mRNP biogenesis, and found that Tho2 plays two roles. In addition to its function as a subunit of the THO complex, upon perturbation of mRNP biogenesis Tho2 targets Rrp6 to chromatin *via* its C-terminal domain. Interestingly, other THO subunits are not enriched on chromatin upon perturbation of mRNP biogenesis and are not necessary for localizing Rrp6 at its target loci. Our study highlights the potential role of Tho2 in co-transcriptional mRNP quality control, which is independent of other THO subunits. Considering that both the THO complex and the RNA exosome are evolutionarily highly conserved, our findings are likely relevant for mRNP surveillance in mammals.

Keywords: Rrp6, Tho2, THO complex, mRNP quality control, ChIP-Seq

Introduction

In eukaryotes, mRNA molecules mature by undergoing several complex and highly coordinated processes. First, RNA polymerase II transcribes protein-coding genes into primary transcript mRNAs (pre-mRNAs), which are then 5'-capped, spliced, 3'-cleaved, polyadenylated, and assembled with various proteins (1). This process leads to the assembly of messenger RNA/ribonucleoprotein particles (mRNPs), which are exported from the nucleus to the cytosol for mRNA translation.

The assembly of mRNPs is guided by the C-terminal domain of RNA polymerase II, which recruits the THO complex. THO is a heterotetramer of Hpr1, Mft1, Tho2 and Thp2 (2) that interacts with Tex1 (3), and recruits the mRNA-associated factors Sub2 and Yra1 at the 3'-ends of genes to form the TREX (TRanscription-EXport) complex (4). Through the RNA-binding protein Gbp2 and the post-transcriptionally modified Hpr1 subunit of THO, TREX interacts with Mex67 to flag mRNPs for export from the nucleus to the cytoplasm (5–8).

THO is crucial for proper mRNP biogenesis. Without it, nuclear aggregates form and several GC-rich genes cannot be appropriately transcribed. Moreover, its absence induces genome instability by promoting abnormally frequent formation of R-loops, especially at CAG repeats (3, 9). Moreover, mRNPs without THO are recognized as

aberrant, retained in the nucleus, and their transcripts are degraded by the nuclear RNA exosome (10).

Aberrant mRNPs are recognized by the mRNP quality control (mRNP QC) system. However, as such mRNPs arise only rarely in wild-type yeast cells, they are mainly studied either by investigating THO/Sub2 mutants or by analyzing specific wild-type transcripts under stress conditions (e.g., *HSP104*) (11–14). Such studies showed that aberrant mRNPs are disposed of by the TRAMP complex (Trf4/5-Air2-Mtr4), and the nuclear RNA exosome-associated 3'-5' exonuclease Rrp6. However, these studies failed to identify specific components of the mRNP QC, i.e., proteins, which recognize aberrant mRNPs and flag them for degradation.

To untangle the mechanism of yeast co-transcriptional mRNP QC, our group previously developed a Rho-based experimental system in *Saccharomyces cerevisiae* (15). This system relies on an inducible nuclear version of the bacterial helicase Rho. This protein is a barrel-shaped nuclear hexameric RNA-dependent helicase and translocase responsible for 50% of transcription termination events in bacteria (16). As the helicase binds RNA, it translocates along the molecule and disrupts established RNA-protein interactions. Our genome-wide *S. cerevisiae* Rho-based studies (17) have shown that Rho targets 20% of yeast mRNAs, most likely at their unstructured C-rich regions, which are reminiscent of bacterial Rho interaction sites. Moreover, using the model *PMA1* locus, we have shown that Rho helicase impairs transcript processing and packaging factor recruitment during early and late transcription (14). Thus, our experimental system generates many nuclear mRNPs that are marked as aberrant by mRNP QC, retained in the nucleus, and subsequently degraded by a pathway involving Rrp6.

In this study, we used the yeast-based Rho system to investigate the role of THO in the co-transcriptional mRNP QC step. Consistent with other mRNP QC studies (17, 18), we performed chromatin immunoprecipitation followed by DNA sequencing (ChIP-Seq) to analyze, among others, actively transcribed and DNA-associated pre-mRNAs (local transcripts). This methodology enabled us to study mRNP QC of a population of nascent, co-transcribed mRNPs. Our results show that Rho excludes Mft1 and Hpr1, while it enriches Tho2 and Rrp6 on chromatin, and that the enrichment of Rrp6 depends on the C-terminal domain (CTD) of Tho2. On this basis, we propose

a model in which Tho2 recognizes aberrant mRNPs, and marks them for degradation by the Rrp6 subunit of the nuclear RNA exosome.

Materials and Methods

Yeast strains and growth conditions

S. cerevisiae strains were derived from the wild-type strain BMA41 (MAT**a** *ade2-1 ura3-1 leu2-3,112 his3-11,15 trp1Δ can1-100*) (19). Deletions and tagging (Table S1) were performed via one-step gene replacement (20) and were verified by PCR. Tagged loci were amplified by PCR and Sanger sequenced, and the presence of the tag additionally verified by immunoblotting. The doxycycline-regulated Rho helicase construct (pCM185-RhoNLS, Tet-off, *TRP1* centromeric plasmid) was described previously (15).

S. cerevisiae was grown in a synthetic complete medium (2% glucose) without tryptophane (-trp), at 25°C, according to standard procedures. The growth was monitored by measuring OD600. Strains carrying pCM185-RhoNLS were grown under repressive conditions (5 µg/ml of doxycycline). The Rho helicase was induced by switching the yeast cells to the -trp medium without doxycycline, at 25°C/16 h.

ChIP and ChIP-Seq libraries

ChIP was performed as in Moreau et al. (17). Shortly, samples were crosslinked, lysed with 1.2 ml of FA140 buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate and 1:100 Protease Inhibitor Cocktail (Promega)), sonicated, and centrifuged (2,500 g). Twenty microliters of recovered supernatant, containing genomic DNA, were used as an input, and the rest was mixed with anti-c-MYC sc-40X antibody (Santa Cruz Biotechnology). Samples were incubated at 4°C/overnight, mixed with protein G PLUS-agarose beads (Sigma), rotated at 4°C/2 h, washed twice with FA140 lysis buffer, twice with FA360 lysis buffer (FA140 buffer with 360 mM NaCl), once with washing buffer 1 (10mM Tris-HCl pH 8, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Sodium deoxycholate), once with washing buffer 2 (10 mM Tris-HCl pH 8, 1 mM EDTA, 1% SDS) and once with TE buffer. DNA was eluted with the elution buffer (50 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS) (twice at 65°C/10 min). Eluates were de-crosslinked (65°C/overnight),

digested with proteinase K, and purified on the DNA extraction columns (ThermoFisher GeneJET PCR Purification Kit). ChIP libraries were prepared with NEBNext Ultra II DNA Library (New England Biolabs), in duplicates, and sequenced using the Illumina ScriptSeq protocol at the I2BC high-throughput sequencing platform (Gif sur Yvette).

qPCR

DNA was purified with the Qiagen PCR cleanup columns. The immunoprecipitated DNA (output) was normalized to a 1:200 dilution of input DNA and quantified by qPCR (LightCycler 480, Roche), using primers specific for the *PMA1* gene (Table S1). After qPCR, values of the immunoprecipitated samples (output) were normalized to the values of the input DNA. Amplifications were for each sample performed in duplicate. The mean values and standard deviations were calculated from three independent experiments.

Data processing

ChIP-seq reads were quality controlled (FastQC 0.11.5), trimmed (Cutadapt 1.15 (21)) and uniquely aligned to *Saccharomyces cerevisiae* genome V64.1.1 (downloaded via SGD) with bwa 0.7.17 (bwa mem -T 30 -t 6) (22). Alignment omitted strand specificity, employed user-defined minimal quality, and was assessed with Samtools 1.7 (23).

The *in vivo* binding profiles were generated with Deeptools 2.5.3 pipeline (24). Shortly, for each input and immunoprecipitated (IP) replicate, the reads were counted, pooled, and normalized with bamCompare. Read ratios (IP/input) were obtained with bigwig Compare, and profile plots with ComputeMatrix and plotProfile. ChIP-seq analysis was conducted with PePr 1.1.24 (25), which can employ differential mode and perform statistical analysis directly on the replicates. We used the differential mode to compare -Rho and +Rho conditions and the native mode to identify protein-DNA interactions. In the differential mode, the only retained regions were those with a P-value <0.05 and matching currently annotated 5752 yeast protein-coding genes (26–28). The circos and beeswarm plots were generated with R 3.6.2, using the OmicCircos (29) and the beeswarm (https://github.com/aroneklund/beeswarm) packages. Snapshots of ChIP-seq peak distributions were visualized with the IGV browser 2.8.10 (30) and custom R scripts.

Statistical Analysis

The correlation between biological replicates was calculated with the Spearman correlation test. The log2 enrichment of chromatin immunoprecipitated protein samples (-Rho vs +Rho) was calculated with the Welch t-test. The statistical difference in the distribution of immunoprecipitated proteins was calculated with the Kolmogorov-Smirnov test implemented in the Genometricorr package (31).

Results

Rho perturbs mRNP biogenesis by altering the recruitment of THO subunits to local transcripts

Our previous investigations with *PMA1* have shown that Rho activity along the nascent transcript interferes with normal recruitment of mRNA processing and packaging factors, yielding mRNPs that are recognized as defective and eliminated by the QC apparatus in a process involving the 3'-5' exonuclease activity of Rrp6 (14, 17, 32). In addition, it was shown that the reduction in mRNA levels was not due to a direct effect of Rho on the transcription elongation complex, which would lead to premature termination of transcription, because no significant reduction in RNAPII occupancy along the PMA1 gene was detected upon expression of Rho (14). Furthermore, we have shown that Rho interferes with proper mRNP maturation by displacing the THO subunit Mft1 from the DNA-associated pre-mRNA transcript (local transcript) of *PMA1* (14). To extend this analysis, in this study we investigated whether Rho affects the retention of the THO subunits Hpr1, Thp2, and Tho2 on this local transcript. For this investigation we used myc-tagged THO subunits (Tho2, Hpr1, Mft1, and Thp2) as the introduction of a Myc tag does not disturb the Rho induced phenotype or functionality of the constructed stains (Figure S1). ChIP analysis of *PMA1* in strains with myc-tagged THO subunits showed that induction of Rho depleted chromatin-bound Hpr1 and Thp2 (Figure 1A). Remarkably, chromatin-bound Tho2 deviated from this pattern, being enriched five-fold upon induction of Rho, indicating a role for Tho2 in the mRNP QC of aberrant mRNP transcripts. Either the absence of whole Tho2p, or C-terminal part of Tho2 restore a certain mRNAs (*IZH4*, *HXT6* and *GAC1*) to a nearly normal level (Fig. S2). ChIP assay coupled with RNase treatment show that Tho2 recruitment does not depend on RNA binding (Figure 1B) what is in accordance with previously

published results (3, 33). However, in the presence of an activated Rho expression system we see a large decrease in Tho2 recruitment to chromatin upon RNase treatment (Figure 1B). To ensure that we did not observe a *PMA1*-specific effect, we sought to extend this analysis to all protein-coding loci. To this end, we performed a ChIP-Seq analysis of myc-tagged Tho2, Hpr1, and Mft1 strains, in the presence or absence of Rho, and combined the data with our previously published Rho-based myc-tagged Rrp6 ChIP dataset (17). For each condition, we sequenced two biological replicates (85% aligned reads, r > 0.98). Without Rho, the *in vivo* binding profiles of Hpr1, Mft1, and Tho2 followed a similar trend (Figure 2A) (34). Consistent with Meinel *et al.* (18), the THO subunits were recruited progressively more towards the 3‘-ends of the local transcripts, peaking around the transcription termination site (TTS). This suggests that the THO subunits were recruited increasingly efficiently as transcription progressed. In the presence of Rho, the *in vivo* binding profiles of the THO subunits were altered (Figure 2A). Hpr1 and Mft1 levels failed to increase toward the 3'-ends of the local transcripts and remained at basal levels. On the contrary, Tho2 bound to the local transcripts almost non-discriminately. Moreover, the Tho2 signal reached a plateau during mid-elongation, and its intensity was comparable to that of the Tho2 peak in cells lacking Rho. Thus, the data suggest that Rho displaces local transcript-bound Hpr1 and Mft1 and increases Tho2 levels.

As it was previously revealed by ChIP analysis of *PMA1*, the Rho effect on mRNP protein composition is mediated by the displacement of the THO subunits Mft1 and Thp2, but not Tho2, with a concomitant enrichment by components of the Rrp6-dependent QC. To explore the general nature of the process, we analyzed the genome-wide association with chromatin in – Rho and + Rho conditions of Tho2, Hpr1, Mft1, and Rrp6 (Figure 2B and Supplementary data). Using differential peak calling, we identified ChIP-Seq peaks of the THO subunits that overlapped protein-coding regions (Figure 2B). Notably, 82% of the peaks overlapped protein-coding genes in -Rho and 70% in +Rho cells (minimal enrichment 1.5-2.0-fold). Among them, we identified 437 Hpr1 and 1954 Mft1 peaks in -Rho cells, and 147 Hpr1 and 0 Mft1 peaks in +Rho cells. Conversely, we identified 982 Tho2 peaks in -Rho cells and 8750 Tho2 peaks in +Rho cells. Similarly, the number of Rrp6 local transcript-associated peaks substantially increased from 111 in -Rho cells to 4973 in +Rho cells. The latter results could be explained by the fact that Tho2 contributes to QC of Rho-affected, aberrant

local transcripts during the recruitment of Rrp6. Using circos plots, we investigated the genome-wide distribution of ChIP-Seq peaks in protein-coding regions (Figure 2C). On these circular plots of the *S. cerevisiae* genome, signal intensity reflected the binding of THO subunits and Rrp6 to chromatin. The circos plots show that Rho-induced depletion of Hpr1/Mft1 and enrichment of Tho2/Rrp6 occurs on a genome-wide scale.

Tho2 contributes to Rrp6 recruitment to chromatin

Our previous work showing that Mft1 is removed upon Rho induction now extends to the THO subunit Hpr1 (14). We suspect that the other subunits Thp2 and Tho2 are likely to be removed along with Hpr1 and Mft1, because THO is known to function as a tetrameric complex. However, the removal of the Tho2 subunit is masked here by the increased recruitment we observed. This unexpected event is accompanied by an increased recruitment of the Rrp6 subunit, suggesting that binding of Tho2 to the aberrant local transcript is the signal that triggers Rrp6 recruitment.

To investigate this hypothesis, we assayed binding of myc-tagged Rrp6 to *PMA1* local transcripts in *tho2* and *mft1* strains (Figure 3A). ChIP analysis showed that Rho expression triggers an increased recruitment of Rrp6 to aberrant local transcripts in the *mft1* background, whereas the intracellular level of the other subunits of THO complex remains unchanged (not shown), an effect that is also present in wt cells. However, in the *tho2* strain, Rrp6 increased recruitment is no longer observed in Rho expressing cells. Therefore, Rrp6-dependent degradation of *PMA1* mRNA in the *tho2* strain is completely attenuated by Rho induction. Taken together, these results indicate that elimination of defective mRNPs cannot occur in the absence of Tho2.

The genome-wide Rrp6 ChIP-Seq analysis extended this result to other protein-coding loci. These loci overlapped 65% of Rrp6 peaks in -Rho and 80% in +Rho cells (enrichment threshold 1.7-3.5-fold). The analysis showed that, compared to wild-type Rho-expressing cells, the binding of Rrp6 was slightly decreased in the *mft1* +Rho strain but completely abolished in the *tho2* +Rho strain (Figure 3B, Figure S3 and Supplementary data). Differential ChIP-Seq peak calling identified 111 Rrp6 peaks in -Rho and 4973 Rrp6 peaks in +Rho wild-type cells, as well as 338 Rrp6 peaks in -Rho and 2832 Rrp6 peaks in +Rho *mft1* cells, all distributed homogenously across the genome (Figure 3C). Strikingly, peak calling identified 333 Rrp6 peaks in -Rho and 0

Rrp6 peaks in +Rho *tho2* cells. Thus, Tho2 appears to play role in recruitment of Rrp6 to chromatin and - by implication - to aberrant local transcripts, which leads to their degradation.

The Tho2 C-terminal domain mediates binding of Rrp6 to chromatin

Interestingly, Tho2 is the only subunit of the THO complex capable of binding nucleic acids and the C-terminal domain (CTD) of this protein plays a key role in this process. Tho2 CTD, which is required for the function but not the stability of the THO complex (3), could therefore anchor Tho2 to the aberrant mRNA before their Rrp6-mediated elimination.

This hypothesis was evaluated by performing ChIP-Seq analysis in Myc-tagged Rrp6 strains carrying either a truncated version of the Tho2 CTD (*tho2Δ*1408-1597) or an allele lacking the entire nucleic acid binding domain (*tho2Δ*1271-1597) (Figure 4A).

Differential ChIP-Seq peak calling identified 111 Rrp6 peaks in -Rho and 4973 Rrp6 peaks in +Rho wild-type cells, whereas in the cases of partially or fully truncated Tho2 CTD variants, a considerable change was observed: in *tho2*Δ1408-1597 cells, 187 Rrp6 peaks in -Rho but only 2905 Rrp6 peaks in +Rho and in *tho2*Δ1271-1597 cells this effect was aggravated since 585 Rrp6 peaks were observed in -Rho and only 25 Rrp6 peaks in +Rho conditions (Figure 4B and Supplementary data). All peaks were distributed homogenously across the genome (Figure 4C, Figure S4). In conclusion, the strain lacking the entire Tho2 CTD almost completely failed to recruit Rrp6 at the genome-wide level, which underlines the importance of Tho2 CTD for the recruitment of Rrp6 to chromatin under Rho induction.

Rho induction enhanced positional correlation between Tho2 and Rrp6

To gain insight into the functional relationship between Rrp6 and Tho2 in the QC, we used our ChIP-Seq approach to study the distribution of Rrp6 and THO subunits proteins on chromatin. We mapped Hpr1, Mft1, Tho2, and Rrp6 ChIP-Seq peaks at confidence level of 0.95, focused on 491 genes whose Rho-affected aberrant local transcripts were degraded by Rrp6 (17), and correlated their positions in -Rho and +Rho cells (Figure 5A). As expected, without Rho, all THO subunits were strongly correlated in position ($r = 0.51$-0.68). Furthermore, no correlation of the Mft1 and Hpr1

with Rrp6 was observed, whereas a slight correlation with Rrp6 (r = 0.26) was observed for Tho2. This trend was enhanced upon Rho expression since the correlation between Tho2 and Rrp6 reaches 0.42. Moreover, 80% of Rrp6 peaks in +Rho cells overlapped with Tho2 peaks (Figure 5B and Figure S5). In this work we demonstrate that Tho2 possesses a yet unknown function important for the Rrp6-dependent co-transcriptional mRNP QC. Because this activity requires Tho2 C-terminal nucleic acids binding domain, we propose that Tho2 binding to nucleic acids flags aberrant nascent mRNP for subsequent Rrp6-dependent degradation. Further studies under conditions where the two players are stabilized would definitively confirm this attractive hypothesis (Figure 5C).

Discussion

The current mRNP QC model suggests that aberrantly processed and packaged mRNPs are recognized as they are being transcribed and then degraded by Rrp6 (11, 13). However, it fails to identify proteins that mark such mRNPs for degradation. In this study, we used *S. cerevisiae* that expressed a bacterial Rho helicase targeted to the nucleus to investigate whether THO participates in the mRNP QC of aberrant mRNP transcripts. We coupled the Rho-perturbation assay with ChIP-seq to quantify the interactions between THO subunits and chromatin, which resulted in identification of THO subunit Tho2 as an important component of mRNP QC.

The ChIP-Seq analysis of myc-tagged THO subunits showed that in the absence of Rho, Hpr1, Mft1, and Tho2 are increasingly recruited towards the 3‘-end of the local transcripts, and become most abundant around the transcription termination sites (TTS). This pattern coincides with the phosphorylation profile of the RNA polymerase II C-terminal domain (CTD) during transcriptional elongation (18). Taken together, these results suggest that as the Ser2 residue becomes more phosphorylated, the polymerase is increasingly able to recruit THO to nascent pre-mRNAs (31). As a result, THO accumulates at the 3' ends of transcribed genes (35).

In our experiments, expression of Rho disrupted the homeostatic pattern of THO binding. Rho displaced chromatin-bound Hpr1 and Mft1 but additionally recruited Tho2 and Rrp6 to chromatin independently of RNA polymerase II Ser2 phosphorylation. We hypothesize that unbound free Tho2 subunit interacts with local transcripts *via* its

nucleic acid-binding C-terminal domain once Rho disrupts THO (36). As Rho strips all nascent transcript-bound proteins throughout the entire length of the local transcript (16), Tho2 can bind anywhere, not only on the local transcript's 3'-end.

We demonstrated that Tho2 plays role in recruitment of Rrp6 to chromatin, thereby positioning it in close proximity to aberrant nascent transcripts, which are then likely degraded by Rrp6 alone (37, 38) or in cooperation with the nuclear RNA exosome. Other proteins, such as Nrd1, Nab3, and Isw1, that are important for processing and export of ncRNA to the nucleus, are involved in tracing aberrant mRNAs (17, 39, 40). Whether Tho2 cooperates with these proteins or functions independently remains unclear.

Our positional ChIP-Seq peak analysis indicates that Tho2 and Rrp6 are co-localized in +Rho and the -Rho cells. Thus, our data are consistent with previous studies (10, 41) and raise the intriguing possibility that Tho2 and Rrp6 can assemble into a complex, either directly or via interaction with one or several adaptor proteins. Moreover, it also implies that THO and the nuclear RNA exosome can interact at protein-coding genes as they do at snoRNA-encoding loci (42). Tho2 can perform these functions by binding Rrp6 directly or by interacting, for example, with Gbp2. THO recruits this protein to pre-mRNAs *via* Tho2's CTD (7), where it monitors splicing and flags aberrantly spliced transcripts for Rrp6-mediated degradation (5). On the other hand, we have previously shown that Rrp6 is recruited to aberrant local transcripts disrupted by Rho *via* a process involving TRAMP and Nrd1 (14, 16, 17, 32). Notably, Nrd1 is recruited to chromatin via RNA polymerase II CTD, in both Rho-dependent and Rho-independent experimental systems (39). Finally, we find the presence of the Rrp6 peaks under homeostatic, Rho- conditions in the *tho2* mutant, which suggests that a Tho2-independent Rrp6-recruiting mechanism exists.

In -Rho cells, the *mft1* strain recruits Rrp6 to three times as many loci as the wild-type strain. We hypothesize that the absence of Mft1 destabilizes THO, which thus dissociates from local transcripts more easily, rendering them aberrant, even in the absence of Rho. Remarkably, such cells would also contain more unbound Tho2, which could recruit Rrp6 onto aberrant transcripts. On the other hand, we noted that *mft1* +Rho cells recruited Rrp6 two twice as many loci as the wild-type strain. However, this finding may be misleading, as it relies on differential ChIP-Seq peak calling.

Specifically, the more Rrp6-bound loci exist in -Rho cells, the harder it is to detect statistically significant enrichment of Rrp6 in +Rho cells, especially if the enrichment factor is weak.

We have also shown that the Tho2 CTD mediates binding of Rrp6 to chromatin (and possibly aberrant local transcripts). Previous studies reported that THO interacts with nucleic acids *via* this domain. As such, this domain is critical for binding of THO to local transcripts but not for its stability (3). From our results, we conclude that partial deletion of the CTD of Tho2 reduces the affinity of Tho2 for nucleic acids, i.e., for aberrant local transcripts, which are therefore less likely to be degraded by Rrp6. In summary, our study reveals a novel function of the Tho2 subunit, which is independent of its role during the co transcriptional loading of the THO complex during elongation on nascent mRNP. Our results provide unprecedented information on the factors and molecular mechanisms that prevent accumulation of deleterious aberrant mRNP. Further characterization of this co-transcriptional mRNP QC pathway will certainly lead to the discovery of new players and novel events necessary to fine-tune the expression/ regulation of specific transcripts according to cells requirements.

Figure Captions

Figure 1. (A) Rho action interferes in the cotranscriptional deposition of THO subunits packaging factors. *PMA1* ChIP of cells carrying myc-tagged THO subunits (Tho2, Hpr1, Mft1, and Thp2). The cells either did (+Rho) or did not (-Rho) express Rho. The panel shows the averages of three experiments, with error bars denoting standard deviations. The top right inset illustrates the hybridization position of the oligonucleotides used for the qPCR (thick black line; coordinates on ChrVII, 481431-481656). (B) Rho action reveals RNase sensitivity of Tho2 recruitment to transcription site. *PMA1* ChIP results for Myc-tagged Tho2 strain harbouring an empty vector (-Rho) or Rho expression vector (+ Rho) with and without RNase treatment. The average result for samples untreated by RNase is set as a reference at 100%. Percentage after RNase treatment was calculated relative to the untreated samples. The panel shows the averages of two experiments, with error bars denoting standard deviations. The top right inset illustrates the hybridization position of the oligonucleotides used for the qPCR (thick black line; coordinates on ChrVII, 481431-481656).

Figure 2. Quantification of chromatin (local transcript)-bound subunits of THO (A) Average *in vivo* binding profiles of Hpr1, Mft1 and Tho2 with (+Rho, green line) and without the Rho helicase (-Rho, blue line), across 5752 yeast protein-coding loci. The y-axis plots the log2 ratio between immunoprecipitated samples (output) and input DNA. (B) Beeswarm plot summarizing the log2 fold-enrichment of Tho2, Hpr1, Mft1, and Rrp6 ChIP-seq peaks. P-values: 0.05>*>0.01>**>0.001>***, NA: not applicable, Nf: number of analyzed protein-coding genes, Np: number of detected peaks. (C) Circos plots showing the genome-wide distribution of ChIP-Seq peaks across the protein-coding loci in cells expressing (+Rho) or not expressing (-Rho) the Rho helicase.

Figure 3. Tho2 recruits Rrp6 to aberrant local transcripts. Assays were performed with wild-type, *tho2*, and *mft1* cells, expressing (+Rho) or not expressing (-Rho) Rho. (A) *PMA1* ChIP of cells carrying myc-tagged Rrp6. The panel shows the averages of three experiments, with error bars denoting standard deviations. The top right inset illustrates the hybridization position of the oligonucleotides used for the qPCR (thick black line; coordinates on ChrVII, 481431-481656). (B) Beeswarm plot summarizing the log2 fold-enrichment of Rrp6 ChIP-Seq peaks. P-values: 0.05>*>0.01>**>0.001>***, NA: not applicable, NS: not significant, Nf: number of analyzed protein-coding genes, Np: number of detected peaks. (C) Circos plots showing the genome-wide distribution of Rrp6 ChIP-Seq peaks across the protein-coding loci.

Figure 4. The Tho2 C-terminal domain recruits Rrp6 to aberrant local transcripts. Assays were performed with wild-type, *tho2Δ1408-1597*, and *tho2Δ1271-1597* cells, expressing (+Rho) or not expressing (-Rho) Rho. (A) Schematic of the Tho2 constructs. Tho2Δ1408-1597: a Tho2 variant with truncated Tho2 C-terminal domain (lacking residues 1408-1597), Tho2Δ1271-1597: a Tho2

variant lacking the entire Tho2 C-terminal domain (lacking residues 1271-1597). (B) Beeswarm plot summarizing the log2 fold-enrichment of Rrp6 ChIP-Seq peaks. P-values: 0.05>*>0.01>**>0.001>***, NA: not applicable, NS: not significant, Nf: number of analyzed protein-coding genes, Np: number of detected peaks. (C) Circos plots showing the genome-wide distribution of Rrp6 ChIP-Seq peaks across the protein-coding loci.

Figure 5. Rho induction enhanced positional correlation between Tho2 and Rrp6 (A) Colocalization of the Tho2, Hpr1, Mft1, and Rrp6 ChIP-Seq peaks estimated through an EDF correlogram showing the positional correlation of each protein pair. NA stands for non-applicable in case not enough peaks were available to calculate the correlation. (B) Venn diagram plotting the overlap of the Tho2 and Rrp6 ChIP-Seq peaks. (C) The drawing depicts the possible dual role of Tho2, either as part of THO or as a factor that flags aberrant mRNPs, targeting them for degradation by Rrp6. Successfully assembled mRNP is bound by THO and exported to the cytoplasm. Conversely, aberrant mRNPs are bound by Tho2, which recruits Rrp6 that then degrades aberrant local transcripts.

Supplementary Figures

Figure S1. Presence of a Myc tag on subunits Tho2 and Hpr1 of THO complex does not influence Rho-induced growth defect, transcript degradation or protein level. (A) Serial dilution test performed on strains with a C-terminal Myc tag on the subunits of THO transformed either with a plasmid without Rho expression or the plasmid expressing Rho. (B) Northern blot analyses of total RNA extracts from strains with Myc-tagged members of THO complex grown in Rho-repressing or inducing conditions. (C) Western blot analyses of whole protein extracts isolated from strains

harboring Myc-tagged THO grown under Rho-inducing or repressing conditions and detected with anti-Myc antibodies (*α-Myc*). The experiment was performed and Rho and Tfs1 were detected.

Figure S2. The histogram shows the level of *IZH4*, *HXT6* and *GAC1* mRNA relative to 18 S rRNA as determined by quantitative RT-PCR and by setting the value of mRNA signal to 1 for the wild-type strain in the absence of Rho expression. The average of three independent experiments is shown, with *error bars* representing S.D.

Figure S3. IGV snapshots of ChIP-Seq peaks of Rrp6 and THO subunits (Tho2, Hpr1, Mft1) in wild-type and tho2 strains over protein-coding loci, in cells expressing or not expressing the Rho helicase. These snapshots illustrate that *tho2* cells mostly fail to recruit Rrp6 to local transcripts. The y axis corresponds to the mapped reads counts. All mapped reads are shown on the right.

Figure S4. IGV snapshots of Rrp6 ChIP-Seq peaks in wild-type, *tho2Δ1408-1597* and *tho2Δ1271-1597* cells over protein-coding loci, in cells expressing or not expressing the Rho helicase. These snapshots illustrate the importance of the Tho2 C-terminal domain in recruiting Rrp6. The y axis corresponds to the mapped reads counts. All mapped reads are shown on the right.

Figure S5. IGV snapshots of ChIP-Seq peaks of Rrp6 and THO subunits (Tho2, Hpr1, Mft1) over protein-coding loci, in cells expressing or not expressing the Rho helicase. These snapshots illustrate altered levels of Rrp6 and THO subunits upon the induction of the Rho helicase and the overlap between Tho2 and Rrp6 ChIP-Seq peaks. The y axis corresponds to the mapped reads counts. All mapped reads are shown on the right.

Funding

This work was supported by the Conseil regional du Centre-Val de Loire, recurrent funding from the CNRS to A.R.R., a doctoral fellowship from the Ecole Doctorale Santé, Sciences Biologiques et Chimie du Vivant of the University of Orléans to V. B., and Croatian Science Foundation Grant UIP-2017-05-4411 to I. S.

Acknowledgements

The authors dedicate this manuscript to A. Rachid Rahmouni, who always provided support and encouragement. Moreover, we thank Damir Baranašić for reading and commenting the manuscript, Mateja Remenarić for discussions during the early steps of this work, Hélène Bénédetti for providing the polyclonal anti-Ylr179c and anti-Tfs1 antibodies and Domenico Libri for providing the TSS sites used in this study.

Declaration of competing interest

The authors declare no competing interests.

Data availability

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. Additionally, accession numbers for deposited ChIP-seq data at array express database is: E-MTAB-12378.

For access, please use the following link: https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-12378?key=ec0c1b92-e71a-40c7-b9cd-b63a4b8b55f2

References

1. Cramer P. 2019. Eukaryotic Transcription Turns 50. Cell 179:808–812.

2. Chávez S, Beilharz T, Rondón AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ, Lithgow T, Aguilera A. 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. EMBO J 19:5824–5834.

3. Peña Á, Gewartowski K, Mroczek S, Cuéllar J, Szykowska A, Prokop A, Czarnocki-Cieciura M, Piwowarski J, Tous C, Aguilera A, Carrascosa JL, Valpuesta JM, Dziembowski A. 2012. Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. EMBO J 31:1605–1616.

4. Stäßer K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondón AG, Aguilera A, Struhl K, Reed R, Hurt E. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. Nature 417:304–308.

5. Hackmann A, Wu H, Schneider UM, Meyer K, Jung K, Krebber H. 2014. Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. Nat Commun 5:3123.

6. Katahira J. 2012. mRNA export and the TREX complex. Biochim Biophys Acta 1819:507–513.

7. Xie Y, Clarke BP, Kim YJ, Ivey AL, Hill PS, Shi Y, Ren Y. 2021. Cryo-EM structure of the yeast TREX complex and coordination with the SR-like protein Gbp2. Elife 10.

8. Bretes H, Rouviere JO, Leger T, Oeffinger M, Devaux F, Doye V, Palancade B. 2014. Sumoylation of the THO complex regulates the biogenesis of a subset of mRNPs. Nucleic Acids Res 42:5343–5358.

9. Brown RE, Su XA, Fair S, Wu K, Verra L, Jong R, Andrykovich K, Freudenreich CH.

THO and TRAMP complexes prevent transcription-replication conflicts, DNA breaks, and CAG repeat contractions Short title: RNA biogenesis pathways prevent TRCs and CAG repeat fragility https://doi.org/10.1101/2021.12.06.471001.

10. Libri D, Dower K, Boulay J, Thomsen R, Rosbash M, Jensen TH. 2002. Interactions between mRNA export commitment, 3’-end quality control, and nuclear degradation. Mol Cell Biol 22:8254–8266.

11. Rougemaille M, Gudipati RK, Olesen JR, Thomsen R, Seraphin B, Libri D, Jensen TH. 2007. Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. EMBO J 26:2317–2326.

12. Schmid M, Jensen TH. 2008. Quality control of mRNP in the nucleus. Chromosoma 117:419–429.

13. Villa T, Rougemaille M, Libri D. 2008. Nuclear quality control of RNA polymerase II ribonucleoproteins in yeast: Tilting the balance to shape the transcriptome. Biochim Biophys Acta - Gene Regul Mech 1779:524–531.

14. Stuparevic I, Mosrin-Huaman C, Hervouet-Coste N, Remenaric M, Rahmouni AR. 2013. Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP). J Biol Chem 288:31816–29.

15. Mosrin-Huaman C, Honorine R, Rahmouni AR. 2009. Expression of bacterial Rho factor in yeast identifies new factors involved in the functional interplay between transcription and mRNP biogenesis. Mol Cell Biol 29:4033–4044.

16. Stuparević I, Moreau K, Rahmouni AR. 2020. Use of bacterial rho helicase to gain new insights into the targeting mechanism of nuclear RNAS by the exosome-associated exoribonuclease RRP6 and its cofactors in yeast. Period Biol 121–122:147–153.

17. Moreau K, Le Dantec A, Mosrin-Huaman C, Bigot Y, Piégu B, Rahmouni AR. 2019. Perturbation of mRNP biogenesis reveals a dynamic landscape of the Rrp6-dependent surveillance machinery trafficking along the yeast genome. RNA Biol 16:879–889.

18. Meinel DM, Burkert-Kautzsch C, Kieser A, O'Duibhir E, Siebert M, Mayer A, Cramer P, Söding J, Holstege FCP, Sträßer K. 2013. Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II. PLoS Genet 9.

19. Baudin-Baillieu A, Guillemet E, Cullin C, Lacroute F. 1997. Construction of a yeast strain deleted for the TRP1 promoter and coding region that enhances the efficiency of the polymerase chain reaction-disruption method. Yeast 13:353–356.

20. Wach A, Brachat A, Pöhlmann R, Philippsen P. 1994. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast 10:1793–1808.

21. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12.

22. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

23. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

24. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44:W160–W165.

25. Zhang Y, Lin YH, Johnson TD, Rozek LS, Sartor MA. 2014. PePr: a peak-calling prioritization pipeline to identify consistent or differential peaks from replicated ChIP-Seq data. Bioinformatics 30:2568–2575.

26. Villa T, Barucco M, Martin-Niclos MJ, Jacquier A, Libri D. 2020. Degradation of Non-coding RNAs Promotes Recycling of Termination Factors at Sites of Transcription. Cell Rep 32.

27. Challal D, Barucco M, Kubik S, Feuerbach F, Candelli T, Geoffroy H, Benaksas C, Shore D, Libri D. 2018. General Regulatory Factors Control the Fidelity of Transcription by Restricting Non-coding and Ectopic Initiation. Mol Cell 72:955-969.e7.

28. Wilkening S, Tekkedil MM, Lin G, Fritsch ES, Wei W, Gagneur J, Lazinski DW, Camilli A, Steinmetz LM. 2013. Genotyping 1000 yeast strains by next-generation sequencing. BMC Genomics 14:90.

29. Hu Y, Yan C, Hsu CH, Chen QR, Niu K, Komatsoulis GA, Meerzaman D. 2014. OmicCircos: A Simple-to-Use R Package for the Circular Visualization of Multidimensional Omics Data. Cancer Inform 13:13–20.

30. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative Genomics Viewer https://doi.org/10.1038/nbt.1754.

31. Favorov A, Mularoni L, Cope LM, Medvedeva Y, Mironov AA, Makeev VJ, Wheelan SJ. 2012. Exploring massive, genome scale datasets with the GenometriCorr package. PLoS Comput Biol 8.

32. Honorine R, Mosrin-Huaman C, Hervouet-Coste N, Libri D, Rahmouni AR. 2011. Nuclear mRNA quality control in yeast is mediated by Nrd1 co-transcriptional recruitment, as revealed by the targeting of Rho-induced aberrant transcripts. Nucleic Acids Res 39:2809–2820.

33. Abruzzi KC, Lacadie S, Rosbash M. 2004. Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J 23:2620–2631.

34. Gómez-González B, García-Rubio M, Bermejo R, Gaillard H, Shirahige K, Marín A, Foiani M, Aguilera A. 2011. Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. EMBO J 30:3106–3119.

35. Luna R, Rondón AG, Pérez-Calero C, Salas-Armenteros I, Aguilera A. 2019. The THO complex as a paradigm for the prevention of cotranscriptional R-Loops. Cold Spring Harb Symp Quant Biol 84:105–114.

36. Hurt E, Luo MJ, Röther S, Reed R, Sträßer K. 2004. Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. Proc Natl Acad Sci U S A 101:1858–1862.

37. Callahan KP, Butler JS. 2008. Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p. Nucleic Acids Res 36:6645–6655.

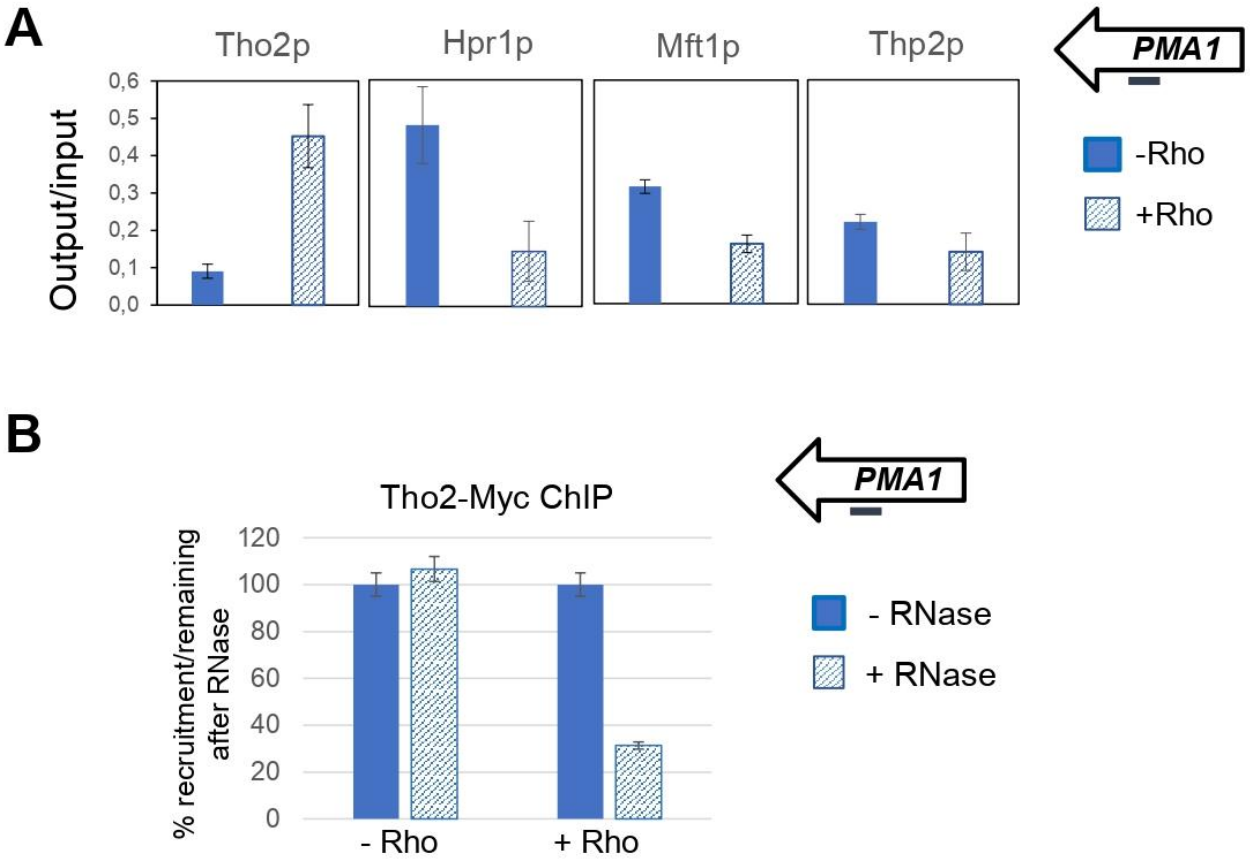
38. Callahan KP, Butler JS. 2010. TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. J Biol Chem 285:3540–3547.

39. Singh P, Chaudhuri A, Banerjea M, Marathe N, Das B. 2021. Nrd1p identifies aberrant and natural exosomal target messages during the nuclear mRNA surveillance in Saccharomyces cerevisiae. Nucleic Acids Res 49:11512–11536.

40. Babour A, Shen Q, Dos-Santos J, Murray S, Gay A, Challal D, Fasken M, Palancade B, Corbett A, Libri D, Mellor J, Dargemont C. 2016. The Chromatin Remodeler ISW1 Is a Quality Control Factor that Surveys Nuclear mRNP Biogenesis. Cell 167:1201-1214.e15.

41. Hieronymus H, Yu MC, Silver PA. 2004. Genome-wide mRNA surveillance is coupled to mRNA export. Genes Dev 18:2652–2662.

42. Larochelle M, Lemay JF, Bachand F. 2012. The THO complex cooperates with the nuclear RNA surveillance machinery to control small nucleolar RNA expression. Nucleic Acids Res 40:10240–10253.


A
Tho2p
Hpr1p
Mft1p
Thp2p
Output/input
0,6
0,5
0,4
0,3
0,2
0,1
0,0
PMA1
-Rho
+Rho
B
Tho2-Myc ChIP
% recruitment/remaining after RNase
120
100
80
60
40
20
0
- Rho
+ Rho
PMA1
- RNase
+ RNase

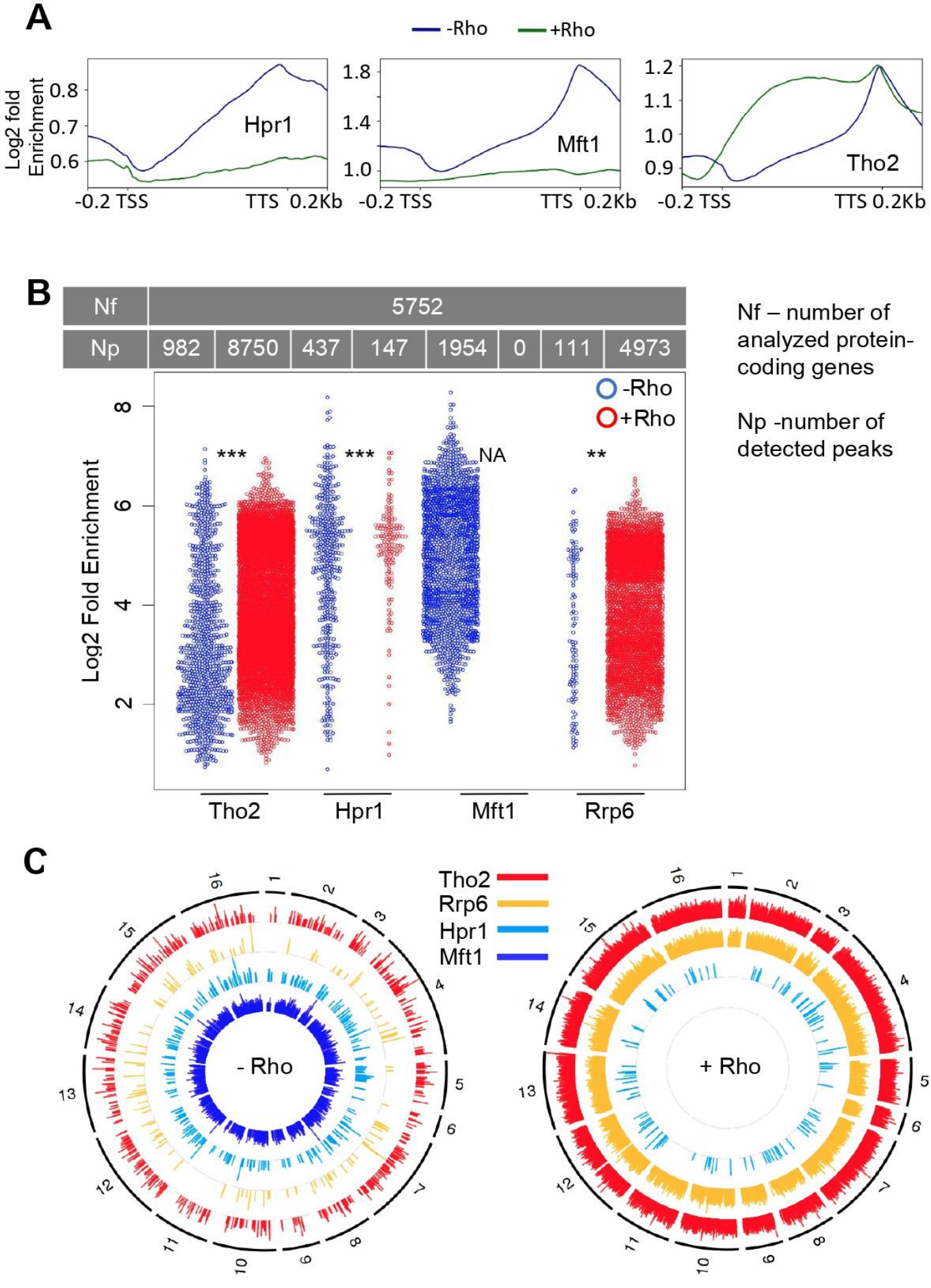
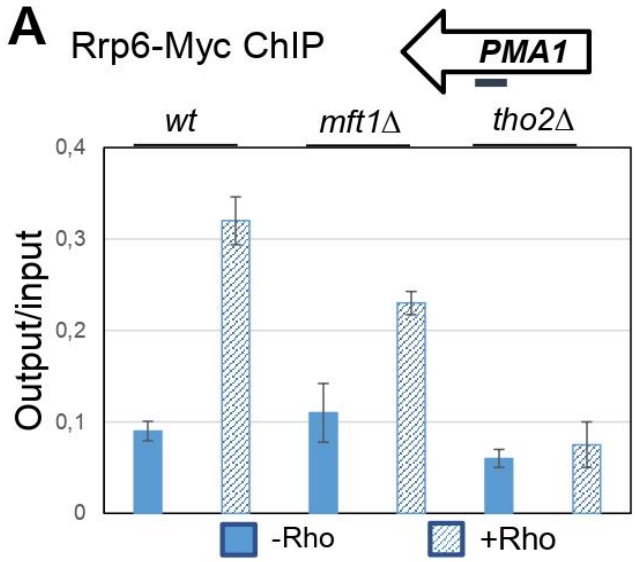
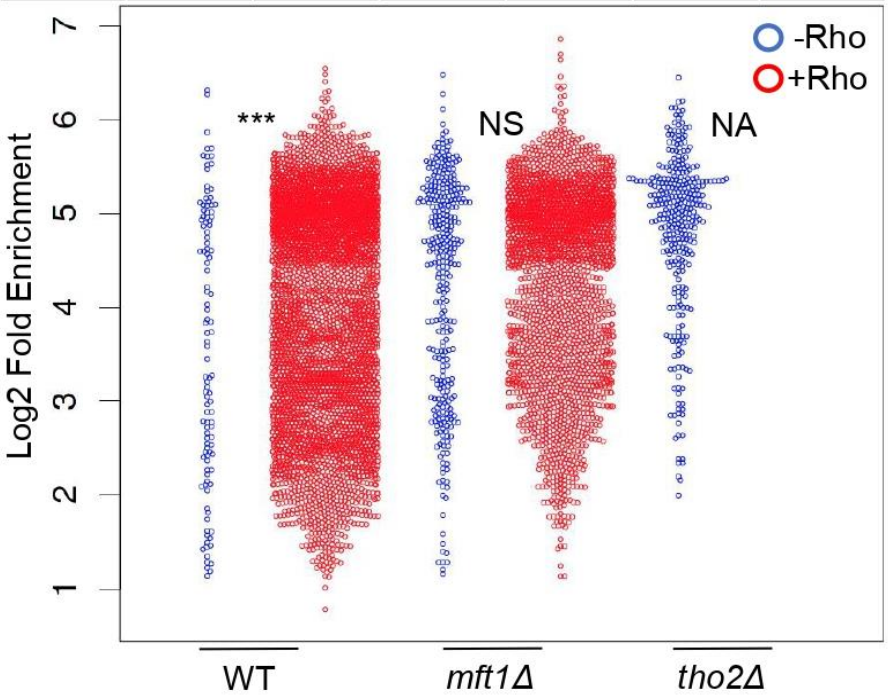

A
-Rho
+Rho
Log2 fold Enrichment
Hpr1
Mft1
Tho2
-0.2 TSS
TTS 0.2Kb
B
Nf
5752
Np
982
8750
437
147
1954
0
111
4973
-Rho
+Rho
Log2 Fold Enrichment
***
***
NA
**
Tho2
Hpr1
Mft1
Rrp6
Nf – number of analyzed protein-coding genes
Np -number of detected peaks
C
Tho2
Rrp6
Hpr1
Mft1
- Rho
+ Rho

Figure 2



B

Nf	5752					
Np	111	4973	338	2832	333	0



Nf – number of analyzed protein-coding genes

Np -number of detected peaks

C

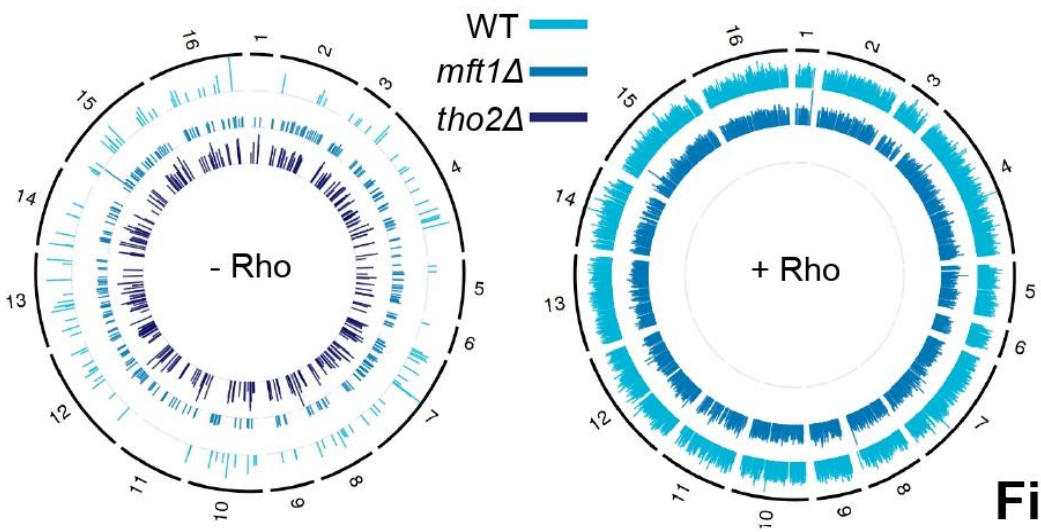


Figure 3

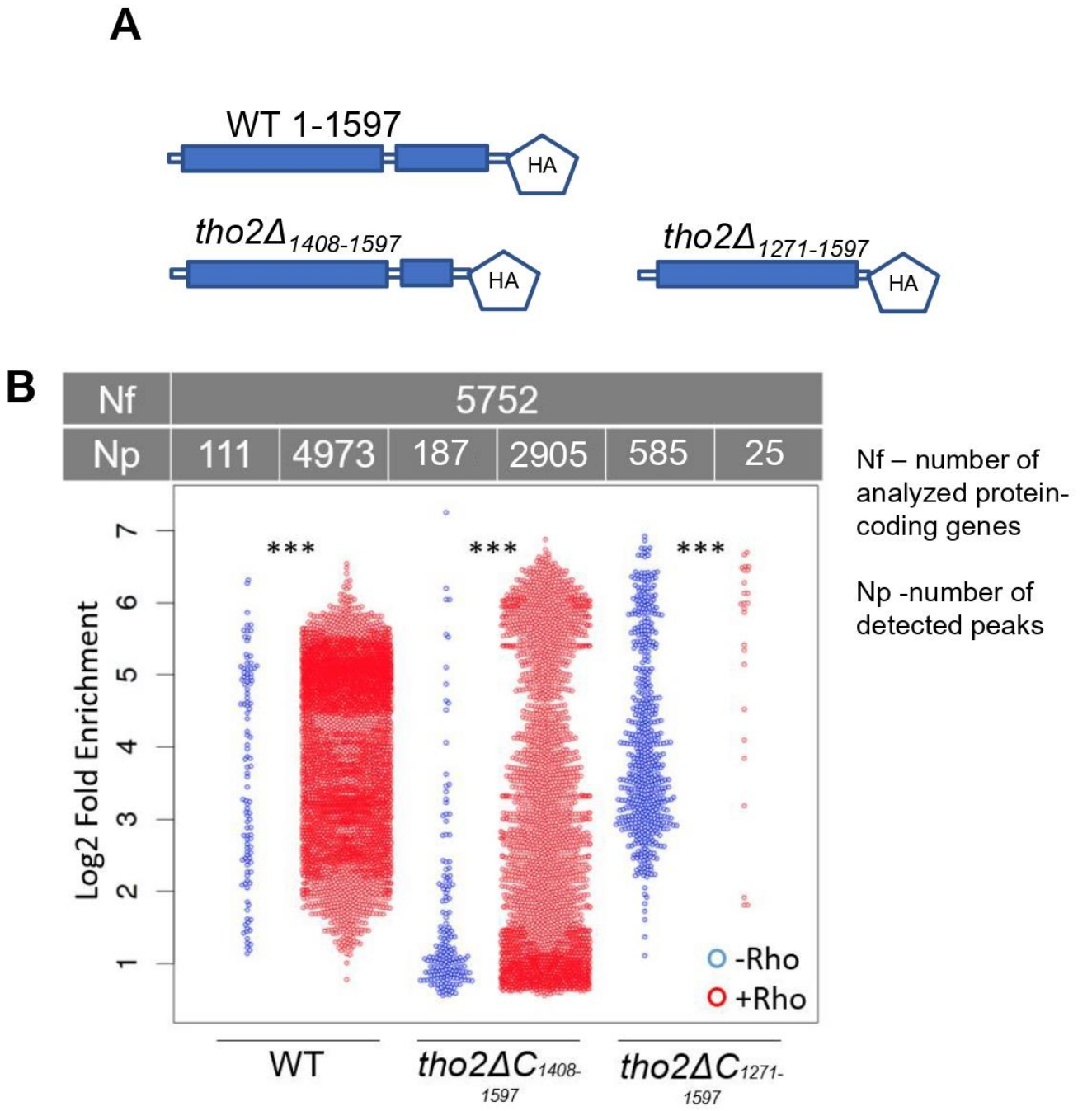

A
WT 1-1597
HA
tho2Δ1408-1597
HA
tho2Δ1271-1597
HA
B
Nf
5752
Np
111
4973
187
2905
585
25
Log2 Fold Enrichment
1
2
3
4
5
6
7
***
***
***
-Rho
+Rho
WT
tho2ΔC1408-1597
tho2ΔC1271-1597
Nf – number of analyzed protein-coding genes
Np -number of detected peaks
C
WT
tho2ΔC1408-1597
tho2ΔC1271-1597
- Rho
+ Rho

Figure 4

Figure S1

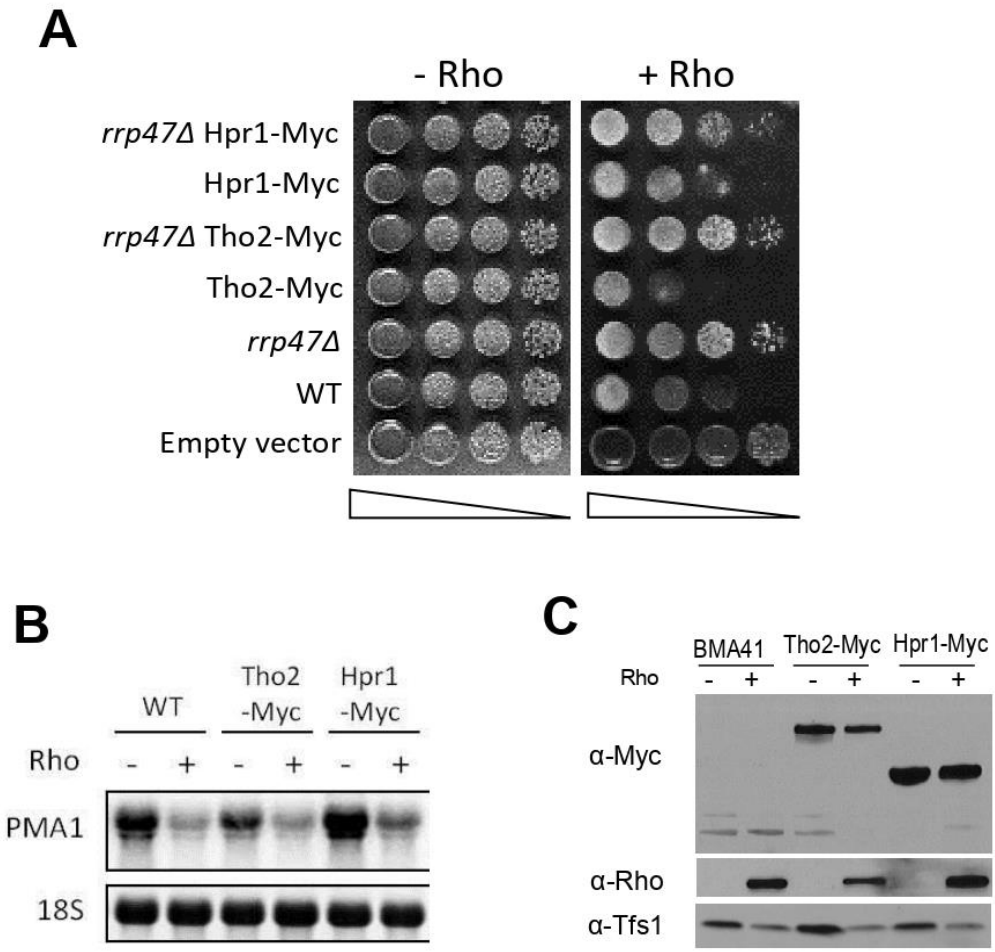

A
- Rho
+ Rho
rrp47Δ Hpr1-Myc
Hpr1-Myc
rrp47Δ Tho2-Myc
Tho2-Myc
rrp47Δ
WT
Empty vector
B
WT
Tho2 -Myc
Hpr1 -Myc
Rho
- + - + - +
PMA1
18S
C
BMA41
Tho2-Myc
Hpr1-Myc
Rho
- + - + - +
α-Myc
α-Rho
α-Tfs1

Figure S2

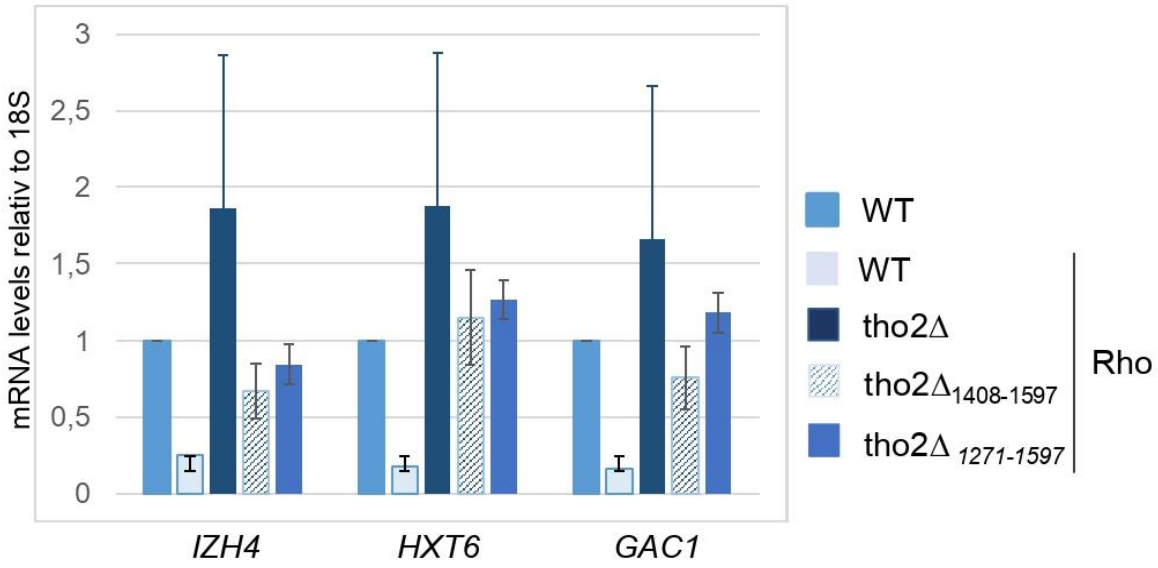
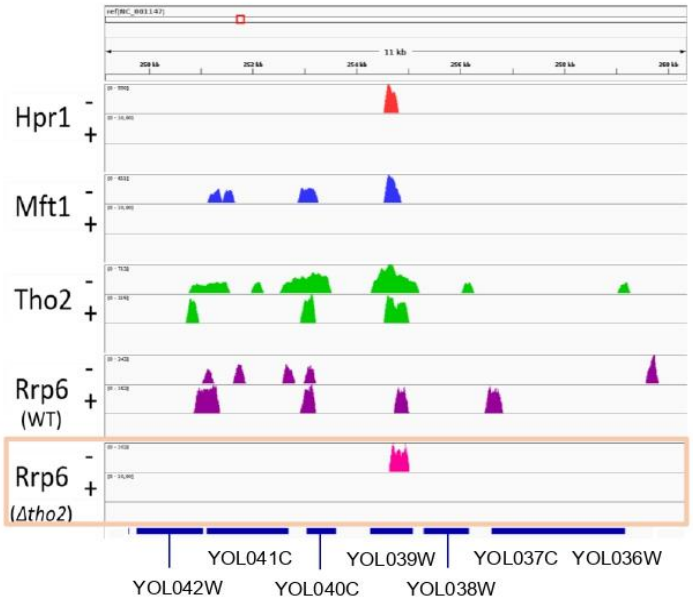
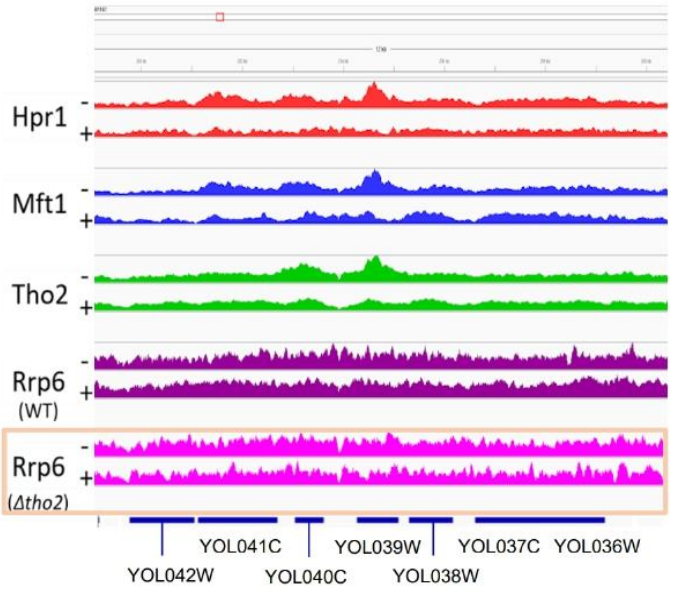
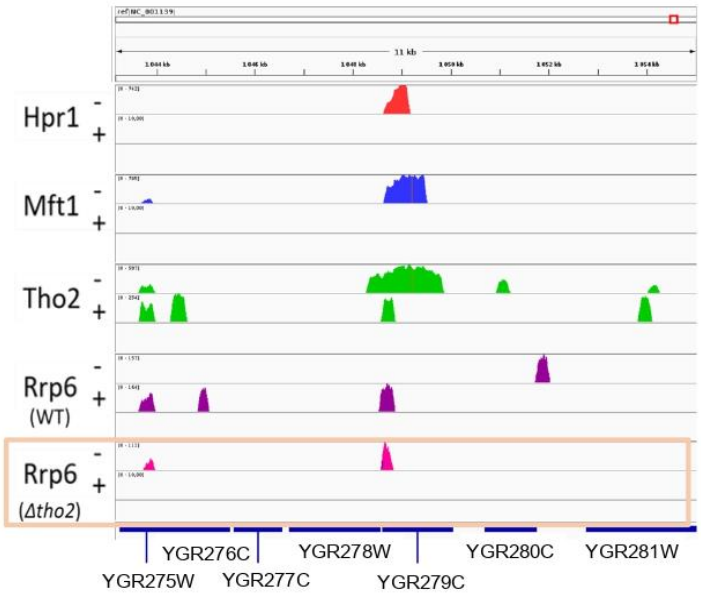
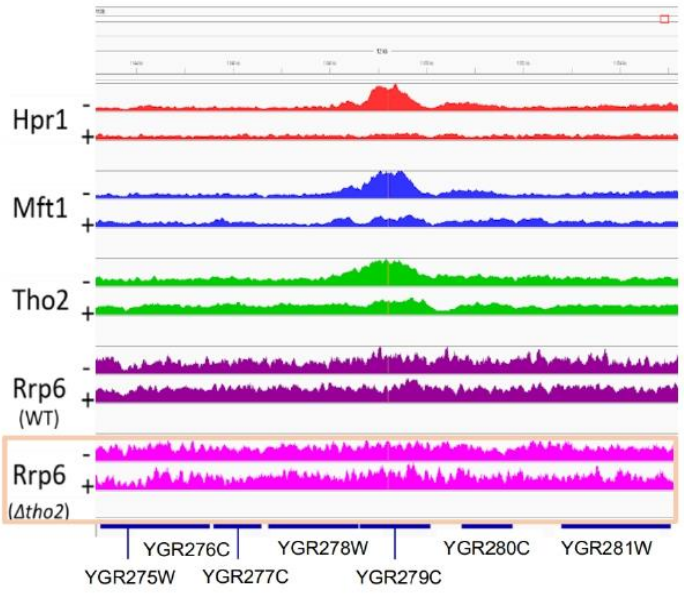

mRNA levels relativ to 18S
3
2,5
2
1,5
1
0,5
0
IZH4
HXT6
GAC1
WT
WT
tho2Δ
tho2Δ1408-1597
tho2Δ 1271-1597
Rho

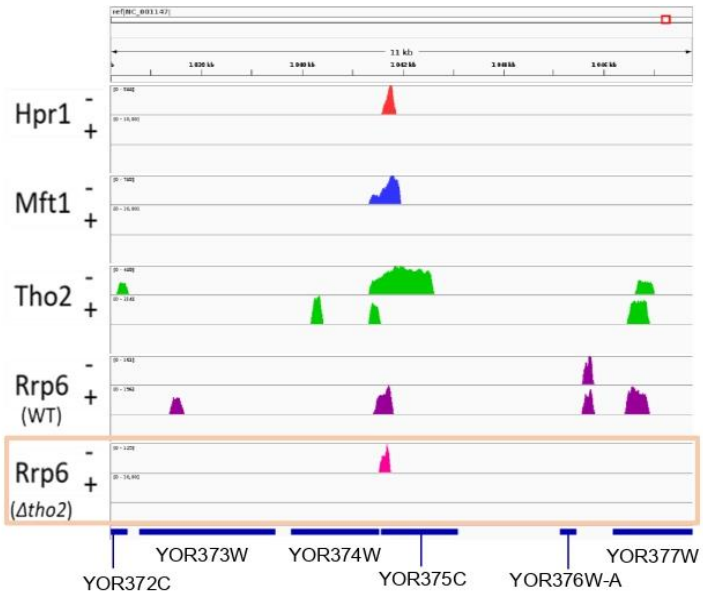
Figure S3


Hpr1
Mft1
Tho2
Rrp6 (WT)
Rrp6 (Δtho2)
YOL042W
YOL041C
YOL040C
YOL039W
YOL038W
YOL037C
YOL036W


Hpr1
Mft1
Tho2
Rrp6 (WT)
Rrp6 (Δtho2)
YOL042W
YOL041C
YOL040C
YOL039W
YOL038W
YOL037C
YOL036W


Hpr1
Mft1
Tho2
Rrp6 (WT)
Rrp6 (Δtho2)
YGR275W
YGR276C
YGR277C
YGR278W
YGR279C
YGR280C
YGR281W


Hpr1
Mft1
Tho2
Rrp6 (WT)
Rrp6 (Δtho2)
YGR275W
YGR276C
YGR277C
YGR278W
YGR279C
YGR280C
YGR281W


Hpr1
Mft1
Tho2
Rrp6 (WT)
Rrp6 (Δtho2)
YOR372C
YOR373W
YOR374W
YOR375C
YOR376W-A
YOR377W

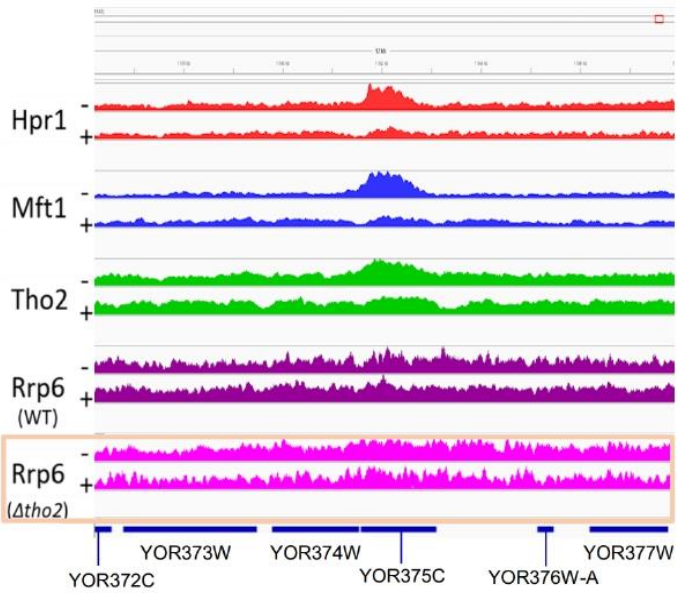

Hpr1
Mft1
Tho2
Rrp6 (WT)
Rrp6 (Δtho2)
YOR372C
YOR373W
YOR374W
YOR375C
YOR376W-A
YOR377W

Figure S4

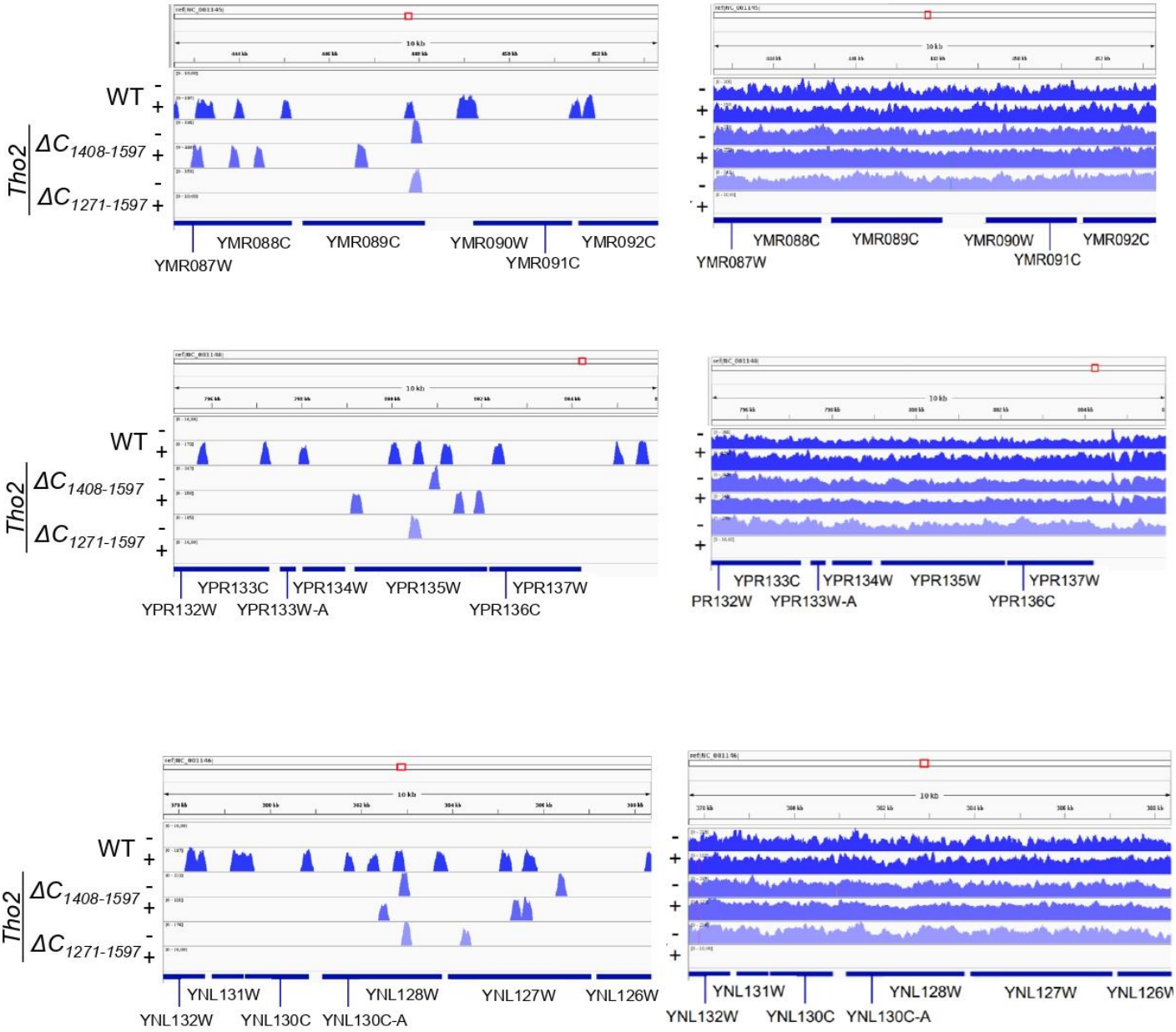
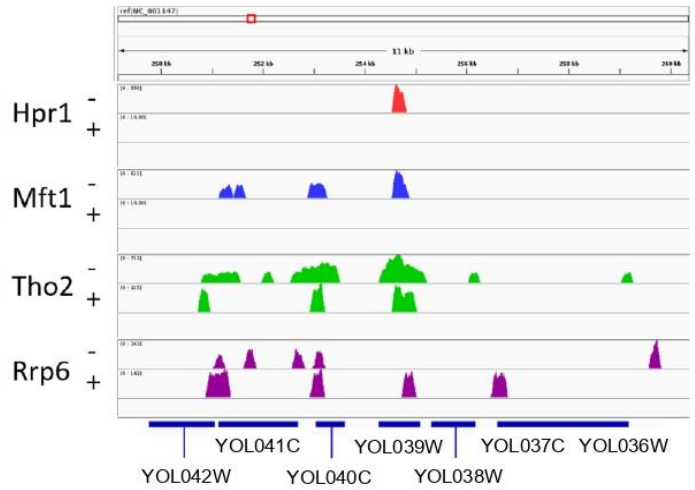
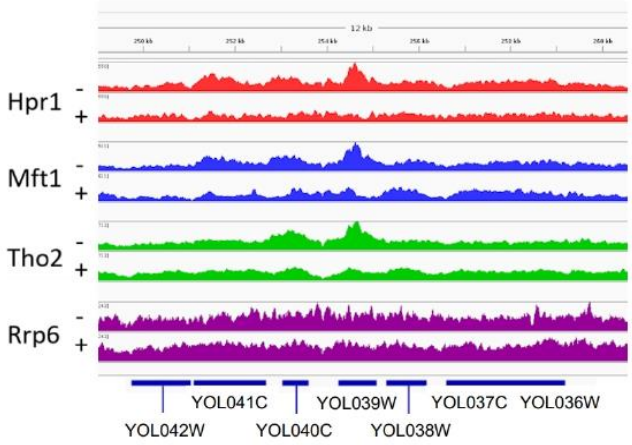
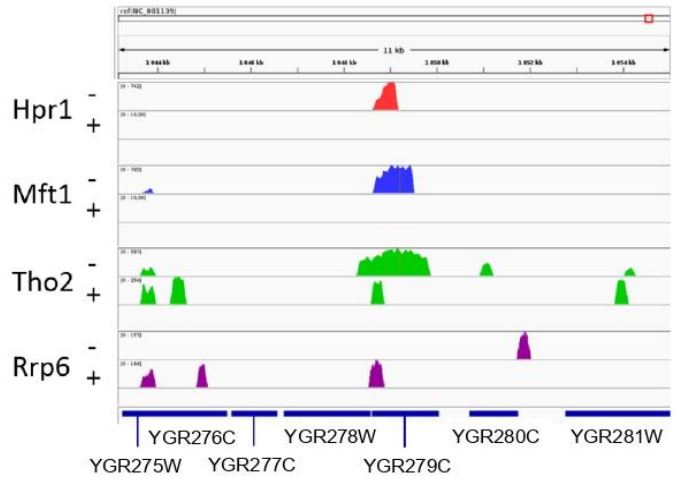
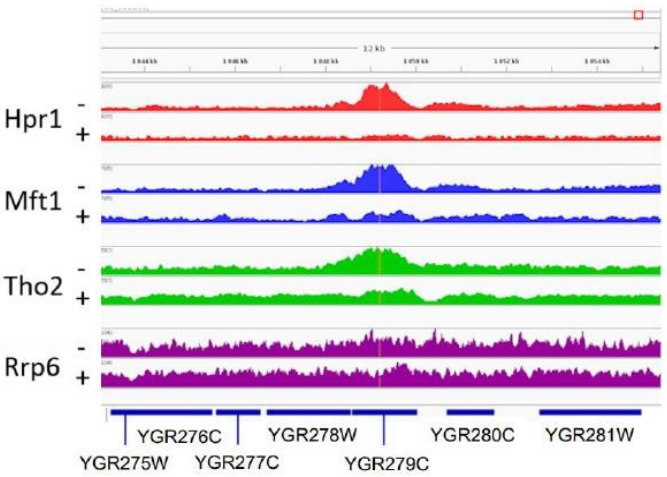

Tho2
WT
ΔC1408-1597
ΔC1271-1597
YMR088C
YMR089C
YMR090W
YMR092C
YMR087W
YMR091C
YPR133C
YPR134W
YPR135W
YPR137W
YPR132W
YPR133W-A
YPR136C
YNL131W
YNL128W
YNL127W
YNL126W
YNL132W
YNL130C
YNL130C-A

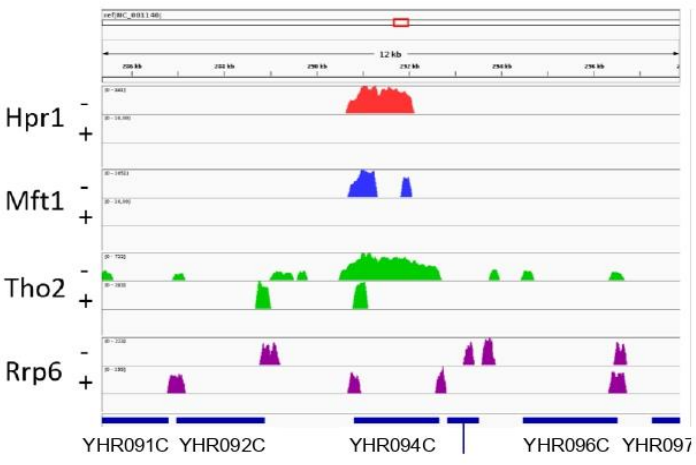
Figure S5


Hpr1
Mft1
Tho2
Rrp6
YOL041C
YOL039W
YOL037C
YOL036W
YOL042W
YOL040C
YOL038W


Hpr1
Mft1
Tho2
Rrp6
YOL041C
YOL039W
YOL037C
YOL036W
YOL042W
YOL040C
YOL038W


Hpr1
Mft1
Tho2
Rrp6
YGR276C
YGR278W
YGR280C
YGR281W
YGR275W
YGR277C
YGR279C


Hpr1
Mft1
Tho2
Rrp6
YGR276C
YGR278W
YGR280C
YGR281W
YGR275W
YGR277C
YGR279C


Hpr1
Mft1
Tho2
Rrp6
YHR091C
YHR092C
YHR094C
YHR096C

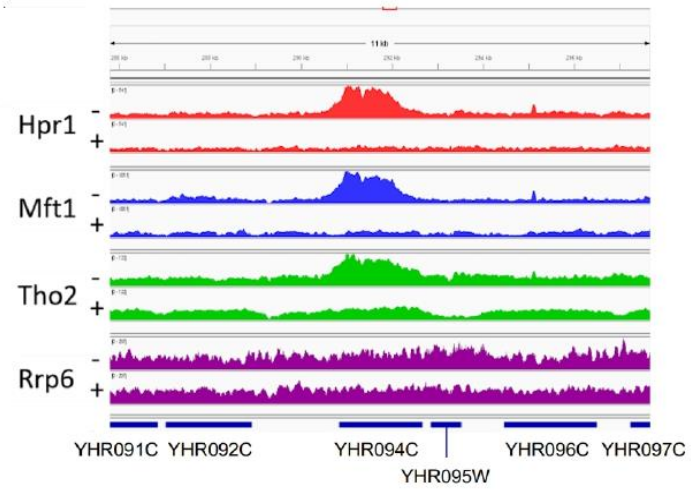
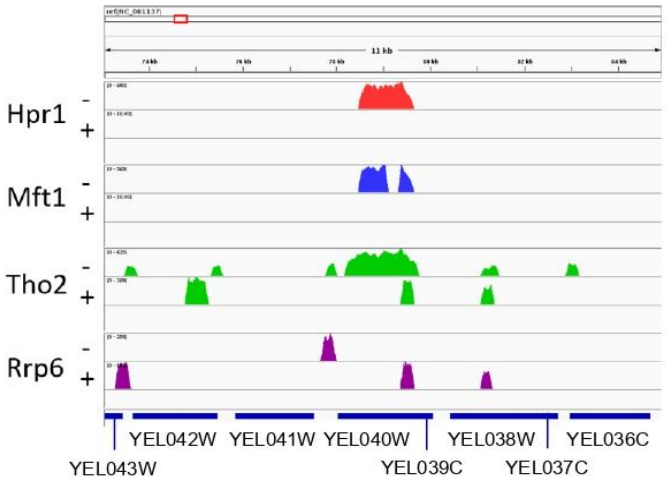
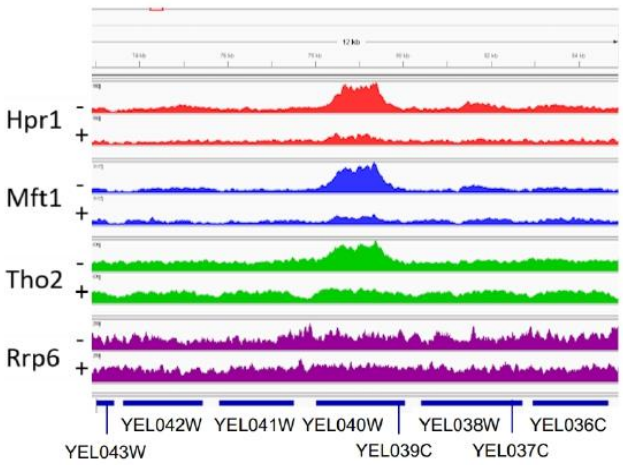
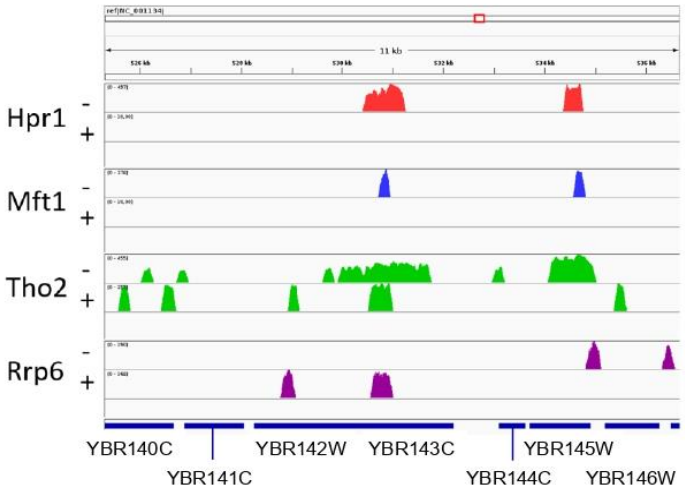
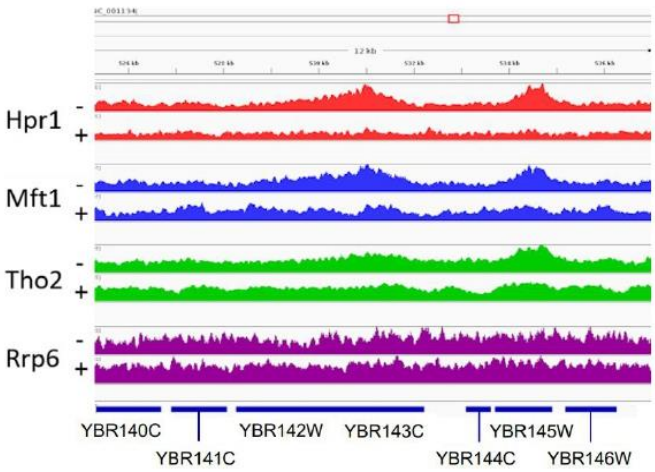

Hpr1
Mft1
Tho2
Rrp6
YHR091C
YHR092C
YHR094C
YHR095W
YHR096C
YHR097C

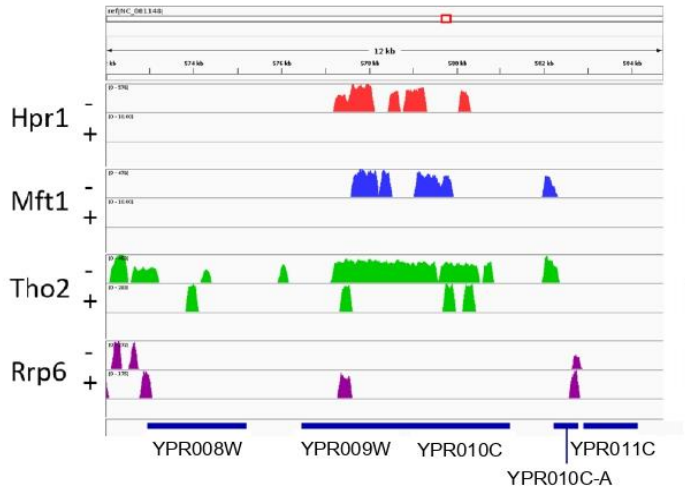
Figure S5


Hpr1
Mft1
Tho2
Rrp6
YEL042W
YEL041W
YEL040W
YEL038W
YEL036C
YEL043W
YEL039C
YEL037C


Hpr1
Mft1
Tho2
Rrp6
YEL042W
YEL041W
YEL040W
YEL038W
YEL036C
YEL043W
YEL039C
YEL037C


Hpr1
Mft1
Tho2
Rrp6
YBR140C
YBR141C
YBR142W
YBR143C
YBR144C
YBR145W
YBR146W


Hpr1
Mft1
Tho2
Rrp6
YBR140C
YBR141C
YBR142W
YBR143C
YBR144C
YBR145W
YBR146W


Hpr1
Mft1
Tho2
Rrp6
YPR008W
YPR009W
YPR010C
YPR011C
YPR010C-A

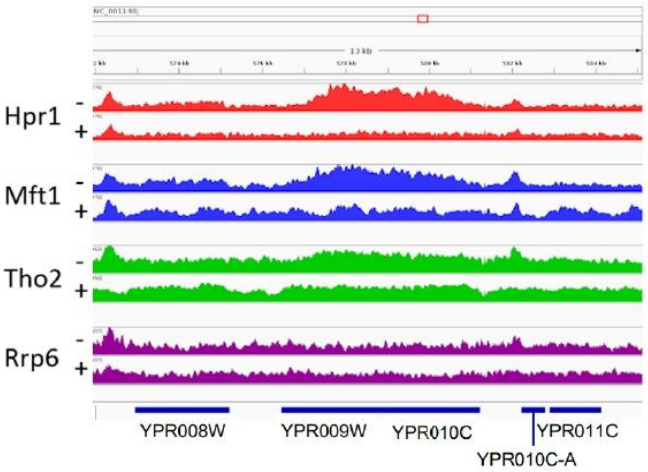

Hpr1
Mft1
Tho2
Rrp6
YPR008W
YPR009W
YPR010C
YPR011C
YPR010C-A

Table S1. Primers

Primer ID	Sequence (5'→3')	Template	Description
tho2::kana S1	GTTGATACATATTCGCACCAGTATACATTTTCAGGACTTTATGC GTACGCTGCAGGTCGAC	pFA6a-kanMX4	Deletion of *THO2*
tho2::kana S2	CAAAGTACACGTTAAAATTCAGCTCGGGTATGTTAAGTACTAGT AATTAATCGATGAATTCGAGCTCG	pFA6a-kanMX4	
hpr1::hph S1	GATGATTTTA ACAATTCAAG AGGCATTAAAACTTGGGCAA AGGAGTAATAATGCGTACGCTGCAGGTCGAC	pFA6a-kanMX4	Deletion of *HPR1*
hpr1::hph S2	ATAGTGGAAA CCGACGCCCC AGCACTCGTC CGAGGGCAAA GGAATAATCTATCGATGAATTCGAGCTCG	pFA6a-kanMX4	
mft1::kana S1	CTGTAAAAAAGGAATCAAAGAACTAAAGCCAAAGGAGACTAA CTCACAATGCGTACGCTGCAGGTCGAC	pFA6a-kanMX4	Deletion of *MFT1*
mft1::kana S2	CTATATGCCTTTTCTATTTAGTAAGAGCTATGCATTATACGTGGT CAATCGATGAATTCGAGCTCG	pFA6a-kanMX4	
tho2Δ$_{1271-1597}$::kana S1	TCGCAAAGACTTCAAAATGATCCACCCAAAAGTGTTGCAAGCG GCAGTGCTGGATTAAATAGCCGCTGCAGGTCGAC	pFA6a-kanMX4	*Deletion tho2Δ$_{1271-1597}$*
tho2Δ$_{1271-1597}$::kana S2	CAAAGTACACGTTAAAATTCAGCTCGGGTATGTTAAGTACTAGT AATTAATCGATGAATTCGAGCTCG	pFA6a-kanMX4	
tho2Δ$_{1408-1597}$::kana S1	AGCTCATATTCTGAAGCTGCTGCTCCTGAATATACCAAAAGATC AAGTAGGTACCGCTGCAGGTCGAC	pFA6a-kanMX4	*Deletion tho2Δ$_{1408-1597}$*
tho2Δ$_{1408-1597}$::kana S2	CAAAGTACACGTTAAAATTCAGCTCGGGTATGTTAAGTACTAGT AATTAATCGATGAATTCGAGCTCG	pFA6a-kanMX4	
THO2-Myc S3	CCGCAAGGTCCCAAGGGTGGGAATTACGTCAGTAGGTACCAG AGGCGTACGCTGCAGGTCGAC	pYM18	Tagging of Tho2
THO2-Myc S2	GTACACGTTAAAATTCAGCTCGGGTATGTTAAGTACTAGTAATT AATCGATGAATTCGAGCTCG	pYM18	
HPR1-Myc S3	GCTACTTCGAACATTTCTAATGGTTCATCTACCCAAGATATGAA ACGTACGCTGCAGGTCGAC	pYM18	Tagging of Hpr1
HPR1-Myc S2	TGCATGAATTTCTTATCAGTTTAAAATTTCTATTAAGAGGATAAT TTAATCGATGAATTCGAGCTCG	pYM18	
MFT1-Myc S3	GGCACTACAAGCGATTTTAGTGCGTCTTCCTCTGTTGAAGAAGT AAAACGTACGCTGCAGGTCGAC	pYM18	Tagging of Mft1
MFT1-Myc S2	CTATATGCCTTTTCTATTTAGTAAGAGCTATGCATTATACGTGGT CAATCGATGAATTCGAGCTCG	pYM18	
THP2-Myc S3	CCATCGGAAAGCTATCCTGTAGATAAAGAAGGTGACATAGTTT TAGAACGTACGCTGCAGGTCGAC	pYM18	Tagging of Thp2
THP2-Myc S2	GTATAACTCTATCTAACTTGTGCAGGCTGGTTAAATAAAATGTG CTTAATCGATGAATTCGAGCTCG		
THO2ΔC12-Myc S3	GCAAAGACTTCAAAATGATCCACCCAAAAGTGTTGCAAGCGGC AGTGCTGGATTAAATCGTACGCTGCAGGTCGAC	pYM18	Tagging of Tho2ΔC$_{1271-1597}$
THO2ΔC12-Myc S2	GTACACGTTAAAATTCAGCTCGGGTATGTTAAGTACTAGTAATT AATCGATGAATTCGAGCTCG	pYM18	
THO2ΔC14-Myc S3	AGATACGACTAGCTCATATTCTGAAGCTGCTGCTCCTGAATATA CCAAAAGATCAAGTAGGCGTACGCTGCAGGTCGAC	pYM18	Tagging of Tho2ΔC$_{1408-1597}$
THO2ΔC14-Myc S2	GTACACGTTAAAATTCAGCTCGGGTATGTTAAGTACTAGTAATT AATCGATGAATTCGAGCTCG	pYM18	
PMA1 F-1010_fwd	GTTTGCCAGCTGTCGTTACCACCAC	-	ChIP qPCR
PMA1 R-1235_rev	GCAGCCAAACAAGCAGTCAACATCAAG	-	

2.2 Yeast RNA exosome activity is necessary for maintaining cell wall stability through proper protein glycosylation

2.2.1 Introduction

The first use of yeasts by humankind dates back to 12000 years ago, when they were used for bread. As humankind evolved, the use of yeasts expanded, both for bread and for alcohol production. More recently, yeasts, and especially *S.Cerevisiae* have been used for production of various compounds via genetic engineering. Considering the position of yeast in these fields, it is not surprising that *S.Cerevisiae* is one of the most widely used model for the study of eukaryotic biology [232]. Interestingly, yeast has also served to study the mRNP degradation pathway, of which the exosome is a key component. More specifically, the exosome harbors two catalytic enzymes Rrp6 and Dis3. Only *S.cerevisiae* mutants harboring a deletion of Rrp6 are used to study how the exosome operates as a deletion of any other exosome subunit, including Dis3, is lethal [148]. Although the deletion of Rrp6 is viable, it leads to two main defects, namely slow growth at physiological temperatures and temperature sensitivity [184, 233]. The involvement of Rrp6 in this physiological feature is not linked to a deficiency or impairment of its catalytic activity, which suggests that Rrp6 has an unexplored, non-catalytic effect [233]. However, the underlying effect leading to these phenotypes needed further investigation.

As Rrp6 degrades a large variety of RNA, we hypothesized that its absence would lead to transcription deregulation that could affect a large variety of metabolic pathways. This hypothesis fits well with a recent finding linking Rrp6-mediated RNA degradation to the cell wall integrity pathway. This pathway is notably used by the cell to ensure cell wall integrity under stress via a MAPK signaling cascade [137]. The addition of an osmotic stabilizer (in our work, sorbitol) during the early phase of the study restored the WT phenotype, confirming a link between Rrp6 and cell wall integrity. Thus, we wondered which cell wall-related genes were affected, i.e. either positively or negatively differently transcribed. To answer this question, I analyzed the expression of cell-wall related genes using the RNA-seq data from [137]. For this analysis, genes were categorized based on their contributions to cell wall integrity. The induction of heat stress by exposure to high temperature (42°C) during 0,10 or 45

minutes in WT and *rrp6Δ* strains gave us insights into the effects of Rrp6 deletion during heat stress on cell-wall related genes. Additional work was performed to determine how the expression of some interesting genes was linked to Rrp6. Rrp6 usually processes and degrades some non-coding transcripts (especially CUTs). As it was shown that promoter sense transcripts and antisense transcripts have regulatory roles in yeast [23, 234-237], I conducted an analysis of the expression of promoter-proximal or antisense ncRNA near cell-wall related genes listed in [137].

2.2.2 Manuscript

Yeast RNA exosome activity is necessary for maintaining cell wall stability through proper protein glycosylation

Ana Novačić[a], Valentin Beauvais[b], Marina Oskomić[a], Lucija Štrbac[a], Aurélia Le Dantec[b], A. Rachid Rahmouni[b], and Igor Stuparević[a,*]

[a]Laboratory of Biochemistry, Department of Chemistry and Biochemistry, Faculty of Food Technology and Biotechnology, University of Zagreb, Zagreb, Croatia; [b]Centre de Biophysique Moléculaire, UPR 4301 du CNRS, 45071 Orléans, France

ABSTRACT **Nuclear RNA exosome is the main 3′→5′ RNA degradation and processing complex in eukaryotic cells and its dysregulation therefore impacts gene expression and viability. In this work we show that RNA exosome activity is necessary for maintaining cell wall stability in yeast *Saccharomyces cerevisiae*. While the essential RNA exosome catalytic subunit Dis3 provides exoribonuclease catalytic activity, the second catalytic subunit Rrp6 has a noncatalytic role in this process. RNA exosome cofactors Rrp47 and Air1/2 are also involved. RNA exosome mutants undergo osmoremedial cell lysis at high temperature or at physiological temperature upon treatment with cell wall stressors. Finally, we show that a defect in protein glycosylation is a major reason for cell wall instability of RNA exosome mutants. Genes encoding enzymes that act in the early steps of the protein glycosylation pathway are down-regulated at high temperature in cells lacking Rrp6 protein or Dis3 exoribonuclease activity and overexpression of the essential enzyme Psa1, that catalyzes synthesis of the mannosylation precursor, suppresses temperature sensitivity and aberrant morphology of these cells. Furthermore, this defect is connected to a temperature-dependent increase in accumulation of noncoding RNAs transcribed from loci of relevant glycosylation-related genes.**

Monitoring Editor
Karsten Weis
ETH Zurich

Received: Aug 20, 2020
Revised: Nov 25, 2020
Accepted: Jan 6, 2021

INTRODUCTION

In eukaryotic cells, 3′→5′ RNA degradation and processing is accomplished through activity of the RNA exosome complex (Chlebowski *et al.*, 2013; Zinder and Lima, 2017; Lingaraju *et al.*, 2020). It plays a major part in RNA metabolism in the nucleus and cytoplasm because it targets almost all RNA classes: its roles include RNA surveillance; mRNA turnover; processing and maturation of rRNAs, snRNAs, and snoRNAs; and degradation of noncoding transcripts (Allmang *et al.*, 1999a; Hilleren *et al.*, 2001; Wyers *et al.*, 2005). It is therefore not surprising that dysregulation of RNA exosome activity broadly impacts gene expression (Van Dijk *et al.*, 2007; Lardenois *et al.*, 2011; Gudipati *et al.*, 2012; Schneider *et al.*, 2012; Bresson *et al.*, 2017; Davidson *et al.*, 2019) and is also implicated in various human malignancies and disorders (Fasken *et al.*, 2020). Rare diseases caused by mutations in genes that encode human exosome subunits (EXOSC proteins) have been termed exosomopathies. They usually encompass single amino acid substitutions rather than more substantial mutations, as RNA exosome activity is essential for viability (de Amorim *et al.*, 2020).

The central part of the highly conserved RNA exosome complex is the exosome core (Exo9). It encompasses nine subunits that form a doughnut-shaped channel that has a structural and regulatory role (Wasmuth and Lima, 2012; Wasmuth *et al.*, 2014). Catalytic activity is provided by two additional subunits: Rrp6, which has exonuclease activity, and Dis3/Rrp44, which has exonuclease and endonuclease

This article was published online ahead of print in MBoC in Press (http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E20-08-0544-T) on January 13, 2021.

The authors declare no conflicts of interest.

*Address correspondence to: Igor Stuparević (istuparevic@pbf.hr).

Abbreviations used: CFW, Calcofluor White; ChIP, chromatin immunoprecipitation; CR, Congo Red; CUTs, cryptic unstable transcripts; CWI, cell wall integrity; EAR, exosome interacting region; ncRNAs, noncoding RNAs; RT-qPCR, reverse-transcription quantitative PCR; TSS, transcription start site.

activities (Briggs *et al.*, 1998; Dziembowski *et al.*, 2007; Lebreton *et al.*, 2008). In yeast, Dis3 is found in both the nuclear and the cytoplasmic isoforms of the exosome complex, whereas Rrp6 is only found in the nuclear isoform, where it additionally associates with its stabilization partner Rrp47 to form the 12-subunit complex Exo-12$^{Dis3/Rrp6/Rrp47}$ (Feigenbutz *et al.*, 2013; Makino *et al.*, 2015). Activity of the nuclear RNA exosome is also stimulated by its cofactors Mpp6 and the TRAMP complex, which function to guide substrate specificity and aid RNA degradation (Schilders *et al.*, 2005; Stuparevic *et al.*, 2013; Wasmuth *et al.*, 2017). The three-subunit TRAMP complex provides RNA-binding (Air1 or Air2), RNA-helicase (Mtr4), and poly(A)-polymerase (Trf4 or Trf5) activities, which play a major role in noncoding RNA degradation (LaCava *et al.*, 2005; Wyers *et al.*, 2005).

All core exosome subunits, as well as catalytic subunit Dis3, are essential in yeast (Mitchell *et al.*, 1997). In contrast, deletion of the gene encoding the catalytic subunit Rrp6 is viable; however, it results in slow growth at physiological temperature and temperature sensitivity (Allmang *et al.*, 1999b; Phillips and Butler, 2003). Interestingly, these two phenotypes are not both caused by the lack of Rrp6 catalytic activity, as Rrp6 catalytic mutants also grow slowly at physiological temperature, but are not temperature sensitive (Phillips and Butler, 2003). Because of that, it has long been clear that Rrp6 has a noncatalytic role in maintaining cell viability upon heat stress, but the molecular nature of this predicament has not been explained. Recent work connected RNA degradation to the cell wall integrity (CWI) pathway, which regulates gene expression to ensure cellular integrity upon stress, through a MAPK signaling cascade (Catala *et al.*, 2012; Wang *et al.*, 2020). Involvement of Rrp6 in this process was inferred from the additive cell wall instability phenotype of *rrp6Δ mpk1Δ* mutant cells, in which CWI signal transduction is inhibited (Wang *et al.*, 2020). Specifically, a role was proposed for a solitary "moonlighting" function of Rrp6, independent of other exosome subunits and its interactors Rrp47 and Isw1, in maintaining CWI at high temperature (Wang *et al.*, 2020).

In this work, we show that the RNA exosome complex is a major regulator of yeast cell wall stability. Exoribonuclease catalytic activity of the Dis3 subunit is essential for maintaining cellular integrity upon heat stress or treatment with cell wall stressors, together with the second catalytic subunit Rrp6 that has a noncatalytic role in this process. The RNA exosome cofactors Rrp47 and Air1/2 also contribute in a significant way. Cells lacking these proteins or Dis3 exoribonuclease activity are not viable at high temperature because of compromised cell wall stability. Importantly, cell bursting and aberrant cell morphology of RNA exosome mutants are suppressed by osmotic support, as well as by overexpression of the Psa1 enzyme, which enables increased production of GDP-mannose that is incorporated into mannoproteins, indicating that protein glycosylation is a major reason for cell wall instability of RNA exosome mutants. Expression of protein glycosylation-related genes *PSA1*, *DPM1*, and *ALG7* is dysregulated in these mutants at high temperature, presumably through mechanisms that involve accumulation of specific noncoding RNAs transcribed from their gene loci.

RESULTS

RNA exosome mutants undergo osmoremedial cell lysis at high temperature

All subunits of yeast RNA exosome complex are essential for viability except for the nuclear-specific catalytic subunit Rrp6, whose inactivation is lethal only above 37°C (Allmang *et al.*, 1999b; Phillips and Butler, 2003). The reason for temperature sensitivity of the *rrp6Δ* mutant remained unknown, especially because Rrp6 catalytic mutants are not temperature sensitive (Phillips and Butler, 2003), and lack of Rrp6 was not linked to significant RNA processing defects at high temperature (Allmang *et al.*, 2000). To test whether *rrp6Δ* cells are inviable at 37°C due to compromised cellular integrity, we supplemented the growth medium with 1 M sorbitol, which acts as osmotic support. We performed all experiments in the W303-derived BMA41 genetic background in which Rrp6-related phenotypes are most pronounced (Klauer and Van Hoof, 2013; Wasmuth and Lima, 2017). Interestingly, osmotic stabilization of the growth medium completely suppressed its temperature-sensitive phenotype and enabled wild-type level of growth after 3 d at 37°C on both YPD and synthetic YNB mediums (Figure 1A). This was due to osmotic stabilization and not sorbitol itself, because the addition of 1 M sucrose, NaCl, or KCl led to a similar level of suppression (Supplemental Figure S1). Also, this effect was not specific to BMA41 genetic background, as growth at 37°C could also be restored with the less temperature-sensitive *rrp6Δ* haploid BY4741 and diploid JHY222 genetic backgrounds (Figure 1B).

Osmotic instability results in cell lysis, so we grew cells in liquid medium for 3 d at 37°C and measured the activity of alkaline phosphatase released into the medium. Alkaline phosphatase is an intracellular enzyme so its release into the medium implies membrane and cell wall lysis. Cells lacking Rrp6 released almost a fivefold higher amount of alkaline phosphatase than wild-type or Rrp6-Y361A catalytic mutant cells, and cell lysis in all strains was completely suppressed by the addition of 1 M sorbitol (Figure 1C) or upon growth at 30°C regardless of sorbitol addition (unpublished data). As higher activity measured with *rrp6Δ* cells could also be due to a change in expression of alkaline phosphatase, we measured intracellular alkaline phosphatase activity with the same cells and found practically no differences between the strains (Supplemental Figure S2), confirming that the extracellular activity observed for the *rrp6Δ* strain is indicative of cell lysis. We also examined the cells by fluorescent microscopy after Calcofluor White (CFW) staining. CFW stains chitin, which in yeast is localized primarily in bud necks and bud scars, as it forms the primary septum (Klis *et al.*, 2002). It was revealed that morphology of *rrp6Δ* cells without osmotic support at 37°C was also consistent with weakened cellular integrity, as the cells were enlarged, unevenly shaped, and grew in clumps (Figure 1D). Based on the intensive staining of cell septa by CFW, it was clearly visible that two or more *rrp6Δ* cells stuck together at their bud necks, meaning the clumps result from a defect in cell separation after division (Figure 1D).

Deletion of the *DIS3* gene encoding the second exosome catalytic subunit is lethal, but it is possible to generate mutants deficient in Dis3 exo- or endoribonuclease activity (Dziembowski *et al.*, 2007; Lebreton *et al.*, 2008). Exo$^-$ (*dis3-D551N*) mutant displays temperature sensitivity (Dra zkowska *et al.*, 2013; Milbury *et al.*, 2019), so we wondered whether the cause is similar as for *rrp6Δ*. Indeed, the addition of 1 M sorbitol restored growth and morphology of this mutant at 37°C and suppressed its cell lysis, as measured by the release of alkaline phosphatase (Figure 2, A–C). Furthermore, we tested viability at 37°C of mutants in monomeric cofactors Mpp6 and Rrp47, as well as viable mutants in subunits of the TRAMP complex, which function as coactivators of the nuclear exosome, and found that the temperature sensitivity and the temperature-induced cell lysis of *air1Δair2Δ* and *rrp47Δ* mutants are also suppressed by osmotic stabilization (Figure 2, D and E). Taken together, temperature-sensitive mutants of RNA exosome catalytic subunits *rrp6Δ* and *dis3 exo$^-$*, as well as mutants in exosome cofactors Rrp47 and Air1/2, undergo osmoremedial cell lysis at 37°C, which is a phenotype indicative of a weakened cell wall.

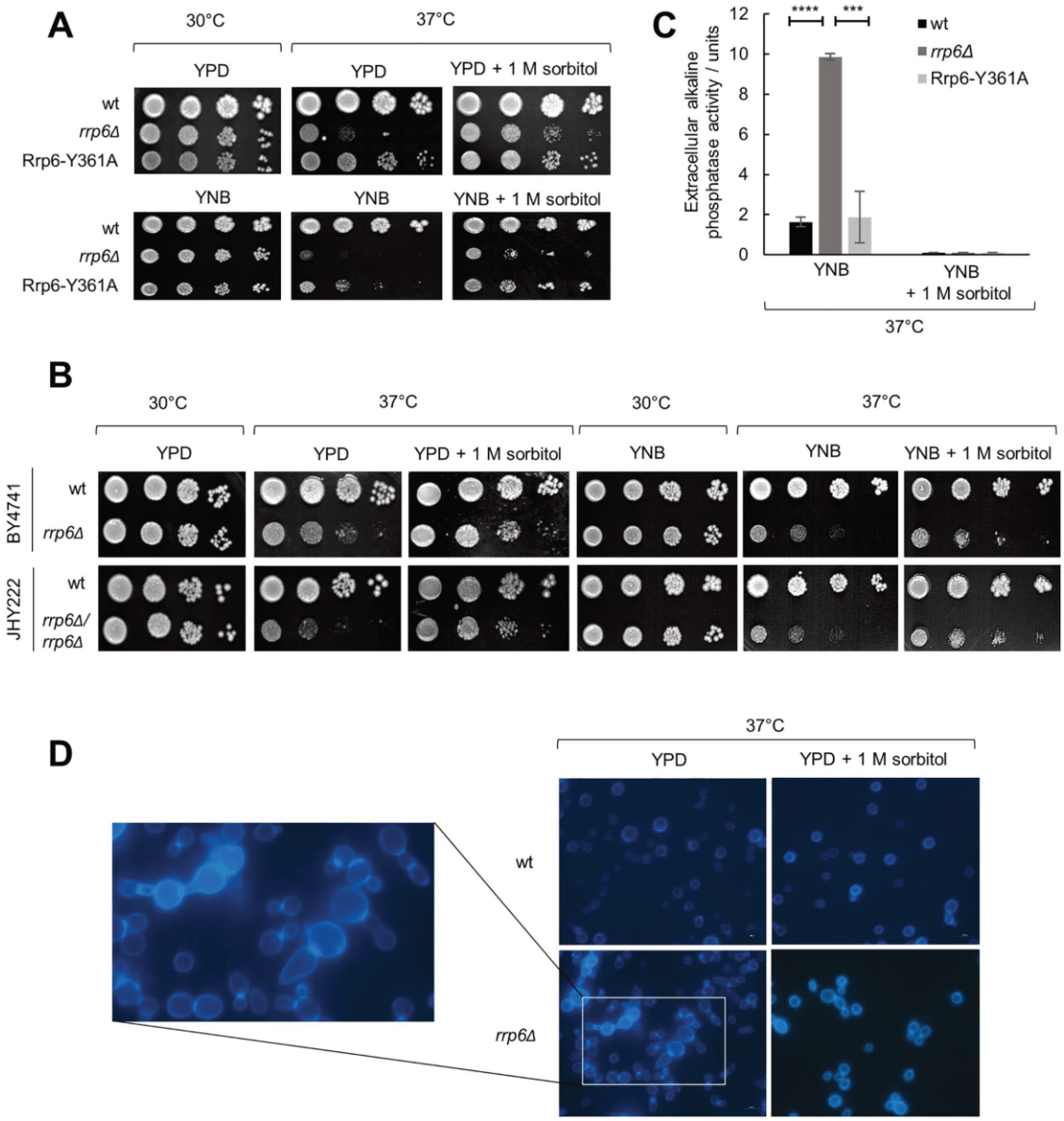


FIGURE 1: Cells lacking Rrp6 display phenotypes indicative of cell wall instability. Strains are BMA41 wild type (wt) and isogenic mutants, unless noted otherwise. Tenfold serial dilutions of cells were spotted on plates and were photographed after 3 d at indicated temperature. (A) Osmotically supporting medium with 1 M sorbitol rescues growth of *rrp6*Δ cells at high temperature. (B) Osmotic support rescues growth of *rrp6*Δ mutants of other genetic backgrounds (haploid BY4741 and diploid JHY222) at high temperature. (C) *rrp6*Δ cells burst at high temperature, unless osmotic support is provided. Strains were grown for 3 d at 37°C and activity of alkaline phosphatase was measured in growth medium. Measurements were performed in duplicate, and reported values represent the means and standard deviations of three independent experiments ($n = 3$). Indicated differences show the significant differences using an unpaired Student's *t* test. Three (***) and four (****) asterisks denote a *p*-value lower than or equal to 0.001 and 0.0001, respectively. (D) Aberrant cellular morphology and cell separation defect of *rrp6*Δ cells at high temperature, visualized by fluorescent microscopy after Calcofluor White staining.

RNA exosome mutants are hypersensitive to cell wall stressors

To investigate whether it is possible to detect cell wall–related phenotypes in RNA exosome mutants at the physiological temperature of 30°C, we examined their growth on media containing known cell wall stressors Congo Red (CR), CFW, caffeine, and SDS. CR and CFW interfere with glucan and chitin assembly, respectively (Roncero and Duran, 1985; Kopecká and Gabriel, 1992); caffeine primarily affects TOR signaling (Kuranda *et al.*, 2006); and SDS is a general cell wall and membrane destabilizer (Popolo *et al.*, 2001). *rrp6*Δ and *dis3 exo*⁻ mutants were hypersensitive to all of these compounds, thereby demonstrating that their cell walls are weaker than those of the corresponding wild-type cells even at the permissive temperature of 30°C when faced with cell wall stressors (Figure 3). Furthermore, their growth was significantly restored by the addition of 1 M sorbitol, which strengthens the argument that the effect is related to cell wall stability (Figure 3).

Regarding mutants in RNA exosome cofactors, for *air1*Δ*air2*Δ and *rrp47*Δ that are temperature sensitive we found that they are also hypersensitive to all tested cell wall stressors, while *mpp6*Δ showed

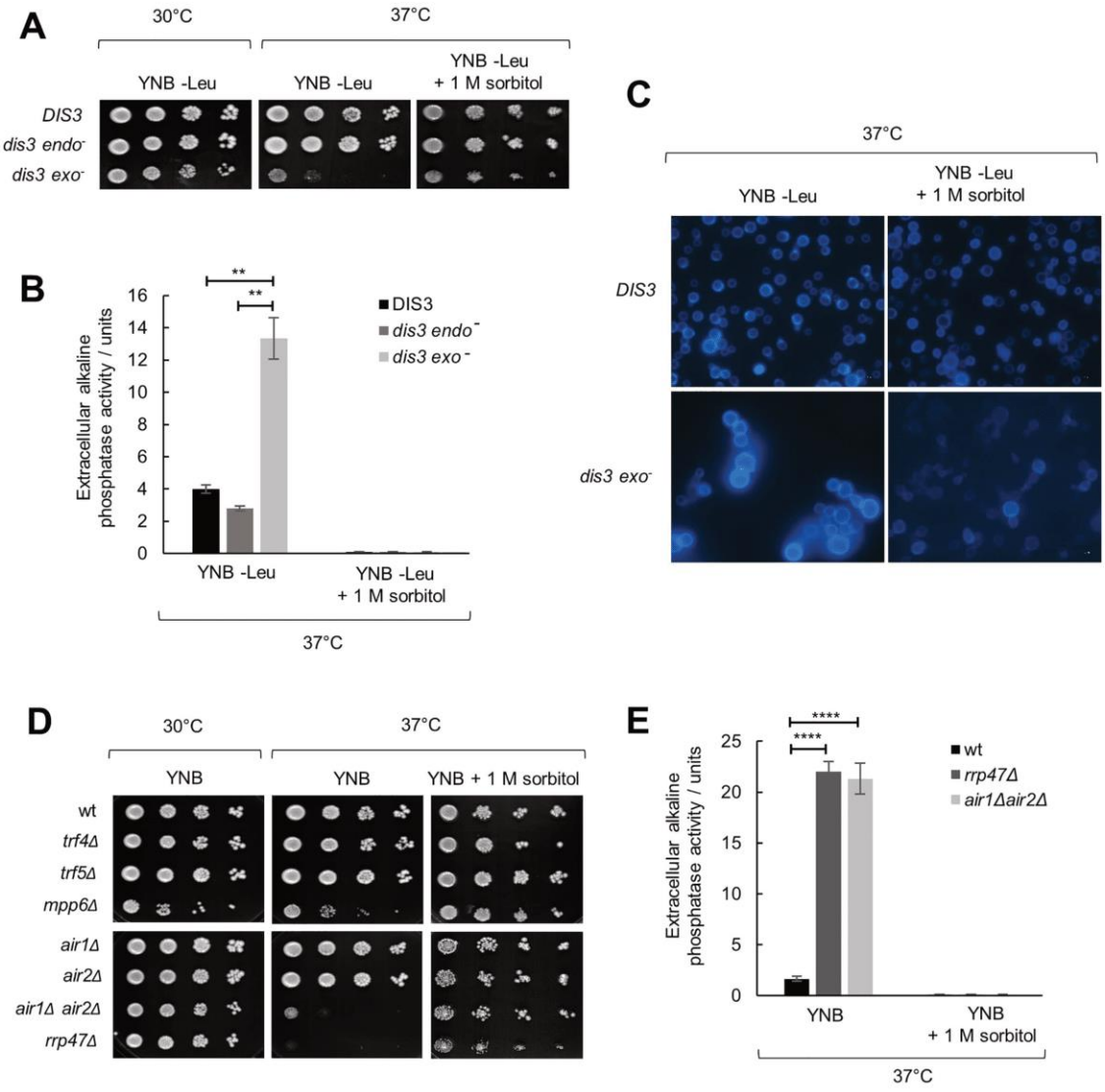


FIGURE 2: Inactivation of Dis3 exonuclease activity and certain exosome cofactors leads to cell wall instability. Strains are W303-derived with genomic copy of *DIS3* gene deleted but bearing a centromeric plasmid that carries the wild-type copy of the *DIS3* gene (*DIS3*) or its alleles with abolished endonuclease (*dis3 endo*⁻, D171N) or exonuclease (*dis3 exo*⁻, D551N) activity. Tenfold serial dilutions of cells were spotted on plates and were photographed after 3 d at indicated temperature. (A) Osmotically supporting medium with 1 M sorbitol rescues growth of *dis3 exo*⁻ at high temperature. (B) *dis3 exo*⁻ cells burst at high temperature, unless osmotic support is provided. Strains were grown for 3 d at 37°C and activity of alkaline phosphatase was measured in growth medium. Measurements were performed in duplicate, and reported values represent the means and standard deviations of three independent experiments ($n = 3$). Indicated differences show the significant differences using an unpaired Student's *t* test. Two (**) asterisks denote a *p*-value lower than or equal to 0.01. (C) Aberrant cellular morphology and cell separation defect of *dis3 exo*⁻ cells at high temperature, visualized by fluorescent microscopy after Calcofluor White staining. (D) Strains are BMA41 wild-type (wt) and isogenic mutants. Mutants in exosome cofactors *rrp47Δ* and *air1Δair2Δ* also show osmoremedial temperature sensitivity. (E) Strains were grown for 3 d at 37°C and activity of alkaline phosphatase was measured in growth medium. Measurements were performed in duplicate, and reported values represent the means and standard deviations of three independent experiments ($n = 3$). Indicated differences show the significant differences using an unpaired Student's *t* test. Four (****) asterisks denote a *p*-value lower than or equal to 0.0001.

specific sensitivity to caffeine (Supplemental Figure S3). The fact that single mutants in the TRAMP RNA-binding subunits *air1Δ* and *air2Δ* did not show cell wall–related phenotypes, but their combined inactivation in the double *air1Δair2Δ* mutant did, indicates their functional redundancy. Inactivation of either Trf4 or Trf5 TRAMP poly(A)-polymerase also did not lead to any cell wall–related phenotypes, while in this case it was not possible to explore whether it is due to functional redundancy because the double mutant is not viable.

Genes involved in protein glycosylation are dysregulated in RNA exosome mutants at high temperature and aiding this process suppresses their temperature sensitivity

The Rrp6-containing RNA exosome is located in the nucleus of the yeast cells, which precludes any direct link to the cell periphery. Instead, given the ubiquitous role of the RNA exosome in gene expression, its role in maintaining cell wall stability upon stress should be visible at the level of mRNAs encoding proteins that are

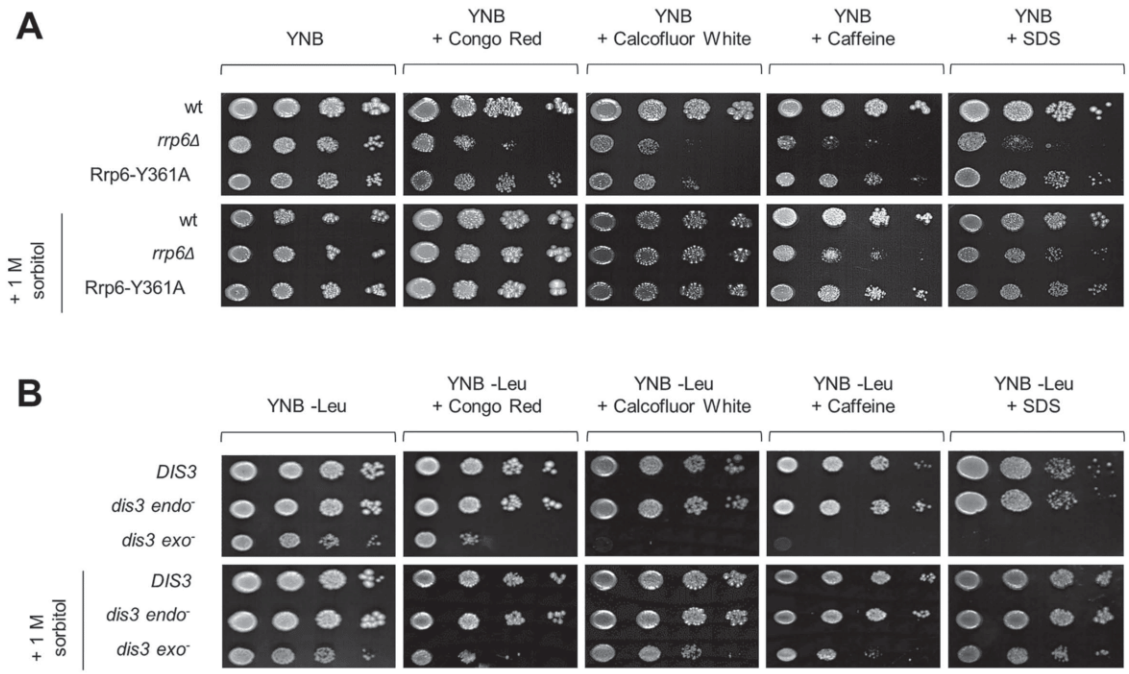


FIGURE 3: Cells lacking Rrp6 protein or Dis3 exoribonuclease activity are hypersensitive to cell wall stressors. Strains are described in Figures 1 and 2. Tenfold serial dilutions of cells were spotted on plates and were photographed after 3 d at 30°C. Concentrations of compounds used: Congo Red 10 µg/ml, Calcofluor White 20 µg/ml, caffeine 6 mM, SDS 0.0075%.

important for cell wall biosynthesis and remodeling. To this aim, we made use of the recently published genome-wide RNA-sequencing analysis that included the datasets of *rrp6*Δ mutant before and after a 45-min heat shock at 42°C (Wang *et al.* 2020). We inspected gene expression profiles of ~180 genes involved in cell wall biogenesis (Orlean, 2012) and visualized them as *rrp6*Δ/wt mRNA ratios on a $\log_2$ scale (Figure 4, A and B, and Supplemental Figure S4). Heat shock–dependent down-regulation in *rrp6*Δ cells as compared with wild-type cells could be seen for a number of genes, such as *GPI12*, encoding an essential protein involved in GPI anchor assembly, and *YPS3*, encoding an aspartic protease (Supplemental Figure S4). However, cell wall–related gene subcategories that encompassed genes with most prominent transcript down-regulation in *rrp6*Δ cells compared with wild-type cells at high temperature were the *precursor supply* gene category, which includes enzymes involved in the synthesis of sugar nucleotides and dolichol phosphate sugars that are precursors for cell wall components, and the *N- and O-glycosylation* category (Figure 4, A and B). In the *precursor supply* gene category, we noticed a strong heat shock–dependent down-regulation of *PSA1* and *DPM1* genes in *rrp6*Δ cells as compared to wild-type cells (Figure 4A). These genes are involved in the synthesis of GDP-mannose and its binding to the dolichol carrier, respectively (Figure 4C). Mannose is exclusively bound to cell wall proteins through N- or O-linked glycosylation in the endoplasmic reticulum and Golgi (Klis *et al.*, 2002). Inspection of the *N- and O-glycosylation* category revealed that *ALG7*, which catalyzes the initial step in synthesis of the oligosaccharide precursor for N-glycosylation (Figure 4C), also showed heat shock–dependent down-regulation in *rrp6*Δ cells compared to wild-type cells (Figure 4B). Even though a large number of genes in this category seemed to be up-regulated in *rrp6*Δ relative to wild-type cells, we hypothesized that protein glycosylation in this mutant should nevertheless be affected, because precursor synthesis and the very early steps in the glycosylation pathway are severely impaired. To experimentally verify whether protein glycosylation is affected in RNA exosome mutants, we analyzed the degree of glycosylation of periplasmic invertase, normally a heavily N-glycosylated protein, by following its electrophoretic mobility with subsequent in-gel activity staining. Periplasmic invertase is easily inducible and is secreted even upon glycosylation defects so it provides a simple readout of the glycosylation status of the cell (Esmon *et al.*, 1987; Belcarz *et al.*, 2002). Positively, we noticed the appearance of a nonglycosylated form of invertase in periplasmic extracts of *rrp6*Δ cells after staining the gel for invertase activity (Figure 4D). This form was also present in periplasmic extracts of other RNA exosome mutants whose cell wall is destabilized: *rrp47*Δ, *air1*Δ*air2*Δ, and *dis3 exo*⁻, and was mostly absent from periplasmic extracts of wild-type and *dis3 endo*⁻ cells (Supplemental Figure S5). Because protein mannosylation is essential for cell viability and its impairment leads to cell wall defects (Janik *et al.*, 2012), this analysis opened the possibility that a general defect in protein glycosylation may be the cause of cell wall instability and therefore temperature sensitivity of RNA exosome mutant cells.

Quantification of *PSA1*, *DPM1*, and *ALG7* mRNAs by reverse-transcription quantitative PCR (RT-qPCR) showed that their levels are lower in *rrp6*Δ, *air1*Δ*air2*Δ, and *dis3 exo*⁻ cells than in corresponding wild-type and *dis3 endo*⁻ cells at high temperature (3 h at 37°C; Figure 5A), in line with their down-regulation observed with *rrp6*Δ cells as compared with wild-type cells upon 45 min of heat shock at 42°C (Figure 4, A and B). For some of these genes, down-regulation in certain RNA exosome mutant cells in comparison to wild-type cells could be observed already at 30°C (Figure 5A), which could explain why glycosylation defects can be detected already at this temperature (Figure 4D and Supplemental Figure S5), even though the effect is not strong enough to cause a detectable phenotype. Out of these three genes, *PSA1* acts most upstream in the protein glycosylation pathway, as it encodes the enzyme GDP-mannose pyrophosphorylase, which synthesizes the activated form of mannose

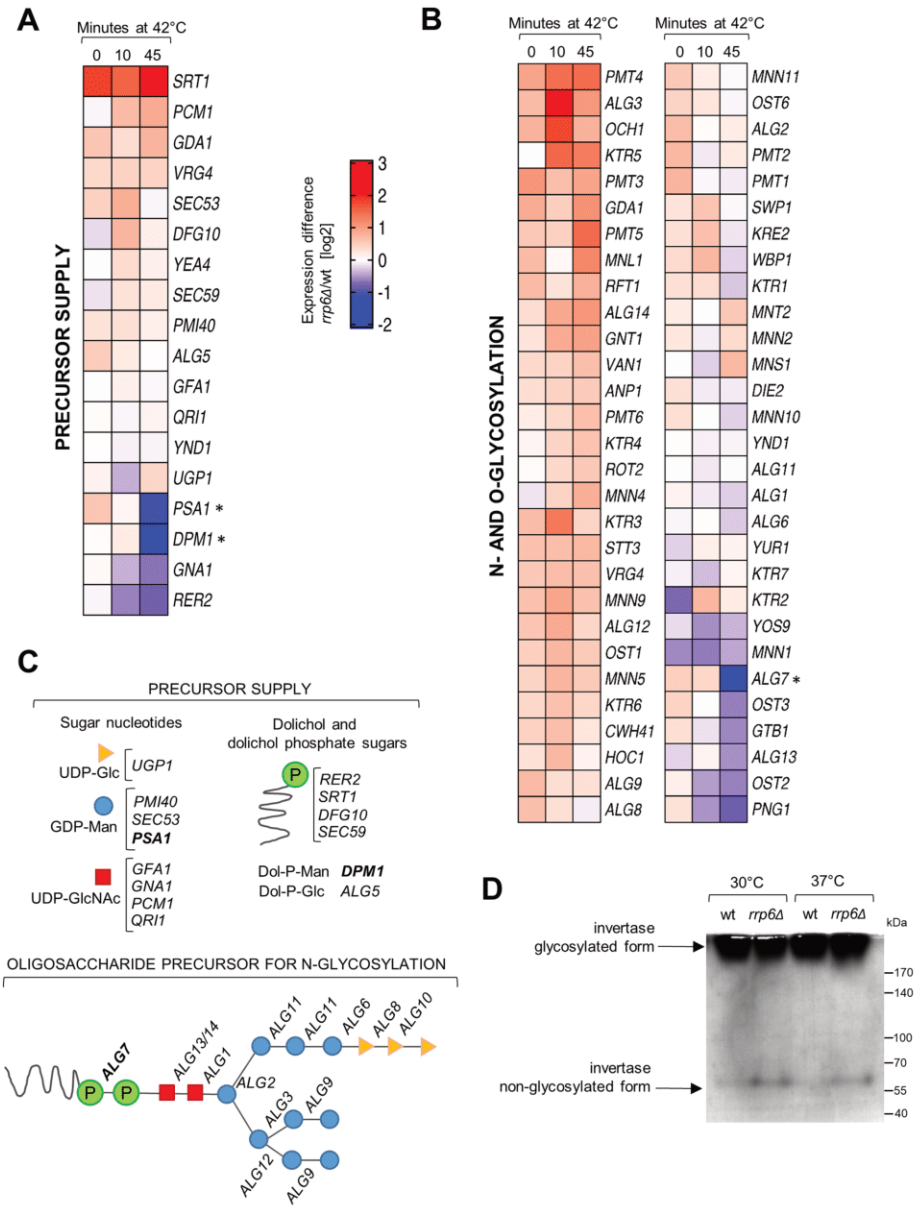


FIGURE 4: Protein glycosylation is dysregulated in cells lacking Rrp6. (A) RNA-seq heat map showing the expression difference of mRNAs encoding genes important for *precursor synthesis* of cell wall components, visualized as *rrp6*Δ/wt mRNA ratio on a $\log_2$ scale. Data is from Wang *et al.* (2020). (B) Same as for A, but for the *N- and O-glycosylation* gene category. (C) Scheme of the genes involved in the synthesis of sugar nucleotides, dolichol and dolichol phosphate sugars, that act as cell wall precursors (above), and in the synthesis of the oligosaccharide precursor for N-glycosylation (below). Genes that are down-regulated in *rrp6*Δ cells upon 45 min at 42°C are marked in bold. (D) Activity staining of invertase from periplasmic extracts. Extracts of *rrp6*Δ cells contain an additional nonglycosylated form of periplasmic invertase, revealing that protein glycosylation is affected in this mutant.

that is incorporated into N- and O-linked glycoproteins (Hashimoto *et al.*, 1997). Psa1 is essential, but partial loss of function of this enzyme or its down-regulation result in phenotypes such as sensitivity to hyposmolarity, cell leakage, and cell separation defects (Zhang *et al.*, 1999; Tomlin *et al.*, 2000; Warit *et al.*, 2000), that are reminiscent of those noticed with the *rrp6*Δ mutant at 37°C. To explore whether the decrease in the *PSA1* mRNA level is reflected by a decrease in the Psa1 protein level in *rrp6*Δ mutant, we C-terminally tagged Psa1 with a Myc tag at its genomic locus and quantified the protein by Western blotting. The Psa1 protein level corresponded well with the mRNA level and confirmed a decrease in the Psa1 level in *rrp6*Δ cells after incubation at 37°C (Figure 5B). Therefore, a low expression level of the essential Psa1 enzyme, or a more general and additive defect in protein glycosylation, could be the reason for the lethality of *rrp6*Δ and other RNA exosome mutant cells at 37°C. To test this hypothesis, we overexpressed Psa1 in these cells from a 2μ plasmid under regulation of its own promoter. Interestingly, overexpression of Psa1 restored the viability of *rrp6*Δ, *rrp47*Δ, and *air1*Δ*air2*Δ mutants at 37°C to a similar degree as osmotic stabilization (Figure 5C), confirming that protein glycosylation was limiting for growth of these cells at 37°C. Overexpression of *PSA1* also partially suppressed the temperature sensitivity of *dis3 exo*⁻ cells (Figure 5C), while in this case the incomplete suppression could be due to the necessity to stably replicate two plasmids (the centromeric plasmid carrying the *DIS3-D551N* allele and the 2μ plasmid carrying the *PSA1* gene) in order to survive. Additionally, overexpression of Psa1 completely suppressed aberrancies in the cell morphology of *rrp6*Δ cells, such as the enlargement of cells and the defect in cell separation (Figure 5D). Taken together, these results demonstrate a role for the RNA exosome in enabling proper protein mannosylation that is needed to preserve cell viability upon temperature-induced stress.

The Rrp6-containing RNA exosome is responsible for the degradation of a class of noncoding RNA transcripts termed CUTs (cryptic unstable transcripts), and inactivation of Rrp6 therefore results in increased CUTs accumulation (Xu *et al.*, 2009). Intriguingly, CUT488 is transcribed in the sense direction through the *PSA1* gene promoter and the 3′ end of this transcript overlaps with the *PSA1* transcription start site (Figure 6A). Quantification of CUT488 by RT-qPCR showed its stabilization in *rrp6*Δ and *dis3 exo*⁻ cells compared with wild-type cells and revealed an additional increase in its level at 37°C (Figure 6B). Because promoter sense transcripts have previously been shown to have gene regulatory roles in yeast (Hainer *et al.*, 2011; Van Werven *et al.*, 2012; Yu *et al.*, 2016), this represents a possible mechanism for *PSA1* down-regulation in *rrp6*Δ and *dis3 exo*⁻ cells. We also found that recruitment of RNA polymerase II to the *PSA1* gene promoter was not significantly changed in these mutant cells compared with wild-type cells at physiological temperature (Supplemental Figure S6). However, at high temperature the occupancy of RNA polymerase II at the *PSA1* gene promoter was drastically decreased in *rrp6*Δ and *dis3 exo*⁻ mutants as compared with wild-type cells or the Rrp6-Y361A catalytic mutant cells (Figure 6C). This was not due to a

general effect on gene transcription in these mutant cells at high temperature, as this effect was not present when probing for RNA polymerase II occupancy at the promoter of the *TAF10* gene (Figure 6D), which is constitutively expressed and does not show any noncoding transcription at its locus. This is conceivably in line with a regulatory mechanism in which accumulation of the normally unstable noncoding RNAs in *rrp6*Δ and *dis3 exo*$^{-}$ cells out-titrates the NNS termination system, thereby promoting read-through of CUT488 into the *PSA1* promoter region, which was recently shown to be a transcriptome-wide phenomenon (Moreau *et al.*, 2019; Villa *et al.*, 2020). Read-through of CUT488 could limit transcription factor and/or RNA polymerase II recruitment to the *PSA1* promoter region and negatively influence transcription of the *PSA1* gene. We also found that the gene loci of the two other down-regulated glycosylation-related genes, *DPM1* and *ALG7*, show transcription of noncoding antisense transcripts at their genomic loci, which are stabilized in *rrp6*Δ cells at high temperature (Supplemental Figure S7). The antisense transcript at the *DPM1* locus was previously mapped as CUT923, while the antisense transcript at the *ALG7* locus was not mapped but can be seen upon inspection of whole-transcriptome tiling array datasets (Xu *et al.*, 2009). Taken together, a possible mechanism for dysregulation of glycosylation-related genes in RNA exosome mutants involves a temperature-dependent increase in accumulation of noncoding transcripts transcribed from their genomic loci.

DISCUSSION

In this work, we demonstrate that the activity of RNA exosome is necessary for maintaining cell wall stability in yeast *Saccharomyces cerevisiae*. RNA exosome mutants undergo osmoremedial cell lysis and show numerous cell wall–related phenotypes that are exacerbated at high temperature. Importantly, this explains that aberrancies in cell wall structure are the reason for temperature sensitivity of these mutants. The essential RNA exosome catalytic subunit Dis3 provides exoribonuclease catalytic activity, while the second catalytic subunit Rrp6 has a noncatalytic role in this process. Besides RNA exosome catalytic subunits, exosome cofactors Rrp47 and Air1/2 are also involved. We show a role for these proteins in maintaining cellular integrity upon heat stress, but also upon treatment with cell wall stressors at physiological temperature, clearly showing that their role is not specific to temperature but to conditions of cell wall stress. Importantly, we provide mechanistic insight into cell wall instability of RNA exosome mutants, as we highlight differential expression of protein glycosylation genes as the factor that disrupts their CWI. Specifically, down-regulation of genes encoding proteins that act in the early steps of the protein mannosylation pathway (*PSA1*, *DPM1*, and *ALG7*) in RNA exosome mutant cells compared with wild-type cells leads to aberrant morphology and temperature sensitivity of these mutants. In addition, artificially aiding protein glycosylation through overexpression of Psa1 suppresses their temperature-sensitive phenotypes, which were previously shown to be due to cell wall instability.

Our results partially contrast with a study that was published during the preparation of this article, which highlighted the role of RNA exosome catalytic subunit Rrp6 in promoting cell survival during heat stress, but argued against involvement of other RNA exosome subunits and cofactors (Wang *et al.*, 2020). They proposed that Rrp6 alone has a highly specialized "moonlighting" function in this process, that is independent of all of its currently known interactors, including its stabilization partner Rrp47 (Wang *et al.*, 2020). Our results clearly show the importance of the essential RNA exosome catalytic subunit Dis3 in this process, as the catalytically inactive *dis3 exo*$^{-}$ (*dis3-D551N*) mutant displays practically identical cell wall aberrancies as *rrp6*Δ mutant (Figures 1–3), which is also the case for mutants in exosome cofactors Rrp47 and Air1/2 (discussed below). This challenges the idea of a highly specialized Rrp6 function in maintaining CWI and is important to delineate, especially as exosome-independent roles of Rrp6 are a highly debated topic in the field (Callahan and Butler, 2008). Furthermore, the potential role of Rrp6 in the CWI pathway was inferred from the additive cell wall instability phenotype of the double mutant *rrp6*Δ *mpk1*Δ, in which a major CWI signaling component was inactivated (Wang *et al.*, 2020). Additivity is suggestive of parallel and redundant functions, and this interpretation was previously applied to the equally severe phenotype of the *rnt1*Δ *mpk1*Δ mutant, which harbors deletion of the dsRNA-specific ribonuclease Rnt1 (Catala *et al.*, 2012). It is, however, clear that Rrp6 has a noncatalytic role in maintaining cellular integrity upon heat stress, as previously implied by the fact that all tested Rrp6 catalytic mutants grow normally at high temperature (Phillips and Butler, 2003). The most straightforward explanation, which fits well with our results, could lie in the well-documented role of Rrp6 in allosterically stimulating the activity of RNA exosome through its C-terminal domain, a process which is independent of Rrp6 catalytic activity (Makino *et al.*, 2015; Wasmuth and Lima, 2017). In line with this, the deletion of only the Rrp6 EAR (exosome-interacting region) domain leads to temperature sensitivity, which pinpoints it as the region of Rrp6 that is necessary for stress resistance (Wasmuth and Lima, 2017).

Besides mutants in the RNA exosome catalytic subunits, we show that for mutants in RNA exosome cofactors temperature sensitivity is also associated with cell wall instability (Figure 2, D and E, and Supplemental Figure S3). Inactivation of the *RRP47* gene, encoding the obligate stabilization partner of Rrp6, results in osmoremedial temperature sensitivity and hypersensitivity to cell wall stressors. Because Rrp47 is critical for Rrp6 protein stability (Feigenbutz *et al.*, 2013; Stuparevic *et al.*, 2013), this result confirms the necessity of Rrp6 protein presence for maintaining cellular integrity upon heat stress. Also, simultaneous inactivation of two homologous genes that encode the TRAMP complex subunits Air1 and Air2 results in cell wall–related phenotypes, in contrast to their individual inactivation. Because Air1 and Air2 function as RNA-binding subunits in different isoforms of the TRAMP complex, this indicates that these isoforms have fully redundant roles in ensuring cellular integrity, which is interesting considering that these isoforms were previously shown to have some nonoverlapping roles based on differential substrate specificity (Schmidt *et al.*, 2012; Stuparevic *et al.*, 2013), somehow similar to what has been recently shown for Trf4 and Trf5 (Delan-Forino *et al.*, 2020). Finally, cell wall instability is an elegant explanation for the observation that *rrp6*Δ phenotype, that is, its temperature sensitivity, is most pronounced in W303 and its derived genetic backgrounds, as wild type of this strain was shown to have an already more destabilized cell wall compared with wild types of other backgrounds (Trachtulcová *et al.*, 2003; Schroeder and Ikui, 2019).

Yeast cell wall is the outermost part of the cell, which determines its shape and provides physical and osmotic protection. It is a polysaccharide network built out of glucan and chitin to which cell wall proteins are bound (Klis *et al.*, 2002). Cell wall proteins function as structural components of the cell wall or enzymes that modify cell wall composition and are often heavily mannosylated through N- or O-linked glycosylation. This modification is vital for yeast, as well as for humans, because it ensures proper protein activity, stability, and localization (Lehle *et al.*, 2006). The essential cytoplasmic enzyme GDP-mannose pyrophosphorylase Psa1 catalyzes the production of

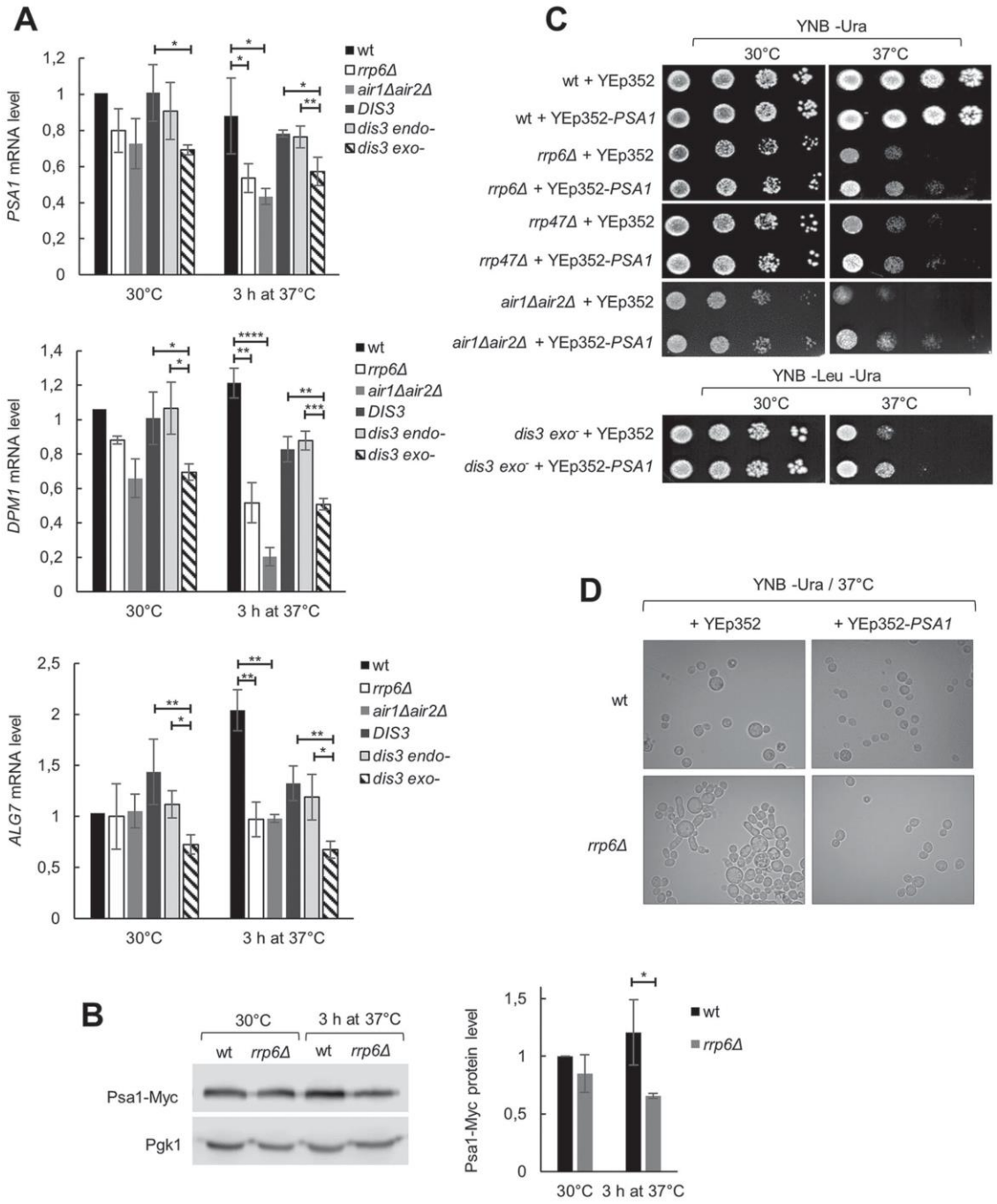


FIGURE 5: Overexpression of *PSA1* rescues temperature sensitivity of RNA exosome mutants. The strains are described in Figures 1 and 2. (A) Levels of *PSA1*, *DPM1*, and *ALG7* mRNAs are lower in *rrp6Δ*, *air1Δair2Δ*, and *dis3 exo⁻* cells than in the corresponding wild-type cells at high temperature. RT-qPCR values are normalized to *PMA1* mRNA and expressed relative to transcript abundance in wild-type cells at 30°C, which is set as 1. Reported values represent the means and standard deviations of three independent experiments ($n = 3$). Indicated differences show the significant differences using an unpaired Student's *t* test. One (*), two (**), three (***), and four (****) asterisks denote a *p*-value lower than or equal to 0.05, 0.01, 0.001, and 0.0001, respectively. (B) Psa1 protein level is lower in *rrp6Δ* than in wild-type cells at high temperature. Myc-tagged Psa1 was quantified by Western blotting. Values are normalized to Pgk1 and expressed relative to protein abundance in wild-type cells at 30°C, which is set as 1. Reported values represent the means and standard deviations of three independent experiments ($n = 3$). Indicated differences show the significant differences using an unpaired Student's *t* test. One (*) asterisk denotes a *p*-value lower than or equal to 0.05. (C) Overexpression of Psa1 from a multicopy plasmid (YEp352-*PSA1*) fully rescues temperature sensitivity of *rrp6Δ*, *rrp47Δ*, and *air1Δair2Δ* cells and partially of *dis3 exo⁻* cells. Tenfold serial dilutions of cells were spotted on plates and were photographed after 3 d at indicated temperature. Control cells were transformed with the empty vector (YEp352). (D) Overexpression of Psa1 from a multicopy plasmid (YEp352-*PSA1*) rescues aberrant phenotype of *rrp6Δ* cells at high temperature.

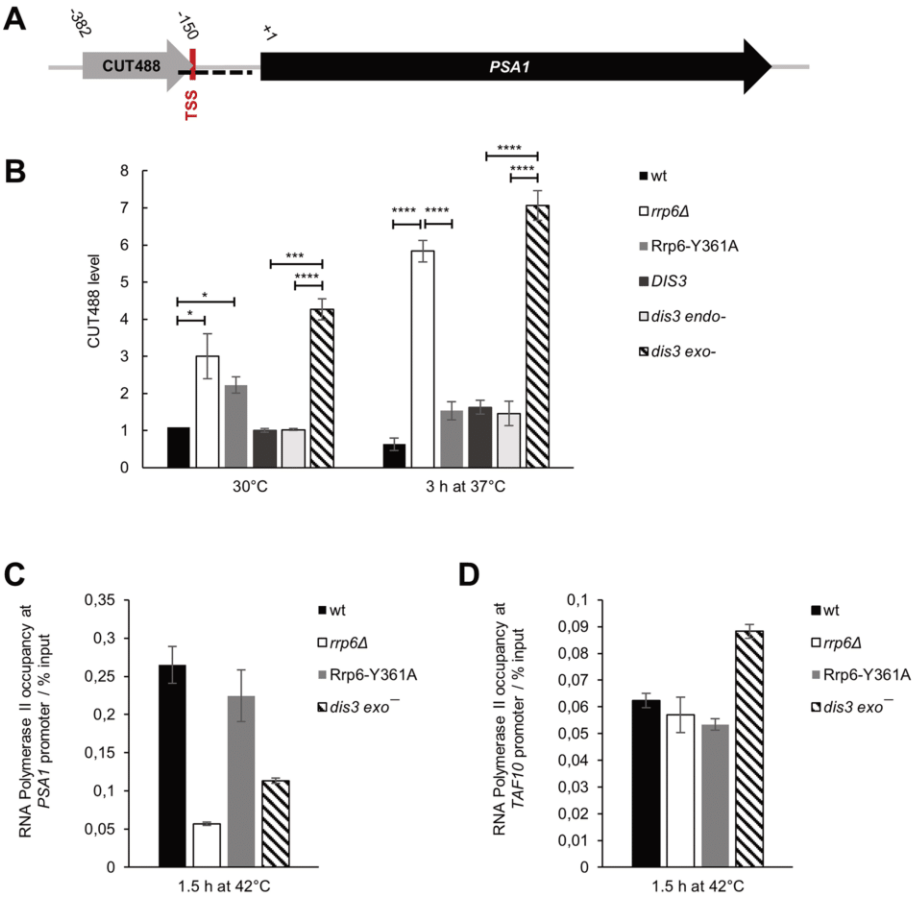


FIGURE 6: Noncoding transcript CUT488 accumulates in cells lacking Rrp6 or Dis3 exoribonuclease activity at high temperature. Strains are described in Figures 1 and 2. (A) Scheme of *PSA1* locus, showing sense transcription of a noncoding transcript CUT488 at its promoter region. Transcription start site (TSS) of *PSA1* is located at position −149 relative to the start of the ORF. The region used for ChIP is marked as a dashed black line. (B) Level of CUT488 RNA is higher in *rrp6*Δ and *dis3 exo*⁻ cells than in corresponding wild-type and Rrp6-Y361A or *dis3 endo*⁻ cells, and that difference is even greater at high temperature. RT-qPCR values are normalized to *PMA1* mRNA and expressed relative to transcript abundance in wild-type cells at 30°C, which is set as 1. Reported values represent the means and standard deviations of three independent experiments ($n = 3$). Indicated differences show the significant differences using an unpaired Student's *t* test. One (*), three (***), and four (****) asterisks denote a *p*-value lower than or equal to 0.05, 0.001, and 0.0001, respectively. (C) Recruitment of RNA polymerase II to *PSA1* gene promoter is decreased in *rrp6*Δ and *dis3 exo*⁻ cells compared with wild-type and Rrp6-Y361A cells at high temperature. Quantification was performed by ChIP of RNA polymerase II using specific antibodies 8WG16. Immunoprecipitated samples (output) were normalized to input following quantification by qPCR. Reported values represent the means and range of two independent experiments ($n = 2$). (D) The decrease of RNA polymerase II occupancy over the PSA1 promoter observed for *rrp6*Δ and *dis3 exo*⁻ cells was not due to a general effect on transcription in these cells because this difference was not present for *TAF10* gene promoter. Quantification was performed as in C.

GDP-mannose, which is the activated form of mannose that gets incorporated into glycoproteins (Hashimoto *et al.*, 1997). Its partial loss of function or down-regulation leads to multiple strong cell wall–related phenotypes such as sorbitol-dependence, cell rupture, and cell separation defects (Zhang *et al.*, 1999; Tomlin *et al.*, 2000; Warit *et al.*, 2000). Importantly, we found that transcription of the *PSA1* gene is down-regulated in RNA exosome mutant cells at high temperature, leading to a lower Psa1 protein level and potentially resulting in *psa1* phenotypes (Figures 4A and 5, A and B). This is strongly supported by rescuing the temperature-sensitive growth and aberrant morphology of these mutants through overexpression of Psa1 (Figure 5, C and D). Because of the previously mentioned strong effects of *PSA1* down-regulation, it is plausible that about a 50% down-regulation of the Psa1 protein level observed in *rrp6*Δ compared with wild-type cells at high temperature can push the protein's enzymatic activity below a physiologically critical level. However, we cannot exclude the possibility that dysregulating expression of genes encoding other proteins involved in the early steps of the glycosylation pathway, such as *DPM1* and *ALG7* (Figures 4, A and B, and 5A) plays a significant contribution in cell wall phenotypes of RNA exosome mutants and that Psa1 overexpression rescues them by generating more precursor supply (Janik *et al.*, 2003). We also hypothesize that a possible reason for transcriptional down-regulation of the *PSA1* gene in RNA exosome mutant cells compared with wild-type cells is the temperature-induced increased accumulation of CUT488, a noncoding transcript transcribed through the *PSA1* gene promoter, which is accompanied by a decrease in RNA polymerase II recruitment at this promoter (Figure 6). At the *PHO84* gene, which is regarded as a model gene for transcriptional regulation through noncoding RNA transcription, loss of Rrp6 was shown to lead to higher production of the antisense transcript due to the decreased recruitment of the NNS complex that normally terminates its transcription (Castelnuovo *et al.*, 2013). Recent transcriptome studies from ours and the D. Libri laboratory showed that out-titration of the NNS complex, accomplished either by perturbation of mRNP biogenesis or inactivation of the RNA exosome, leads to termination defects at ncRNA-producing targets (Moreau *et al.*, 2019; Villa *et al.*, 2020). In line with this mechanism, out-titration of the NNS complex in *rrp6*Δ and *dis3 exo*⁻ cells, because of the more prominent accumulation of ncRNAs in these mutant cells at high temperature, could lead to transcriptional readthrough of CUT488 through the *PSA1* gene promoter and negatively influence *PSA1* gene transcription. Notable examples of loci regulated by noncoding promoter transcription in yeast include *SER3*, *HO*, and *IME1* genes, which are all negatively regulated by transcription of a sense transcript at their promoter regions (Winston *et al.*, 2005; Hainer *et al.*, 2011; Van Werven *et al.*, 2012; Yu *et al.*, 2016). While these are nonessential genes expressed specifically under a certain physiological or life/cell cycle condition, *PSA1* is essential and constitutively expressed. However, it is strongly cell cycle regulated, peaking at the START phase of the cell cycle (Benton *et al.*, 1996), so the possibility of a cell cycle–based regulation of its transcription through noncoding RNA transcription could be an exciting subject for future

investigation. Of course, such broad dysregulation of cell wall stability in *rrp6*Δ cells is most probably due to effects on expression of multiple cell wall–related genes and could involve regulatory roles of noncoding transcription (Novačić *et al.*, 2020), as well as the CWI pathway (Wang *et al.*, 2020). In line with this, we also noticed increased accumulation of noncoding antisense transcripts that are transcribed at *DPM1* and *ALG7* gene loci in *rrp6*Δ cells at high temperature (Supplemental Figure S7).

Another interesting point is that the increase in accumulation of CUT488 at high temperature is independent of Rrp6 catalytic activity but is dependent on the presence of Rrp6 protein and the exoribonuclease activity of Dis3 (Figure 6B). This implies that this CUT is degraded primarily by Dis3 at high temperature and that Rrp6 provides a noncatalytic function in this process, probably that of an equivalent of an RNA exosome cofactor (as discussed in the second paragraph of the *Discussion*). Allosteric stimulation of Dis3 activity by Rrp6 probably happens by direct RNA binding, as well as the widening of the RNA exosome channel through which RNAs need to be threaded to reach the active site of Dis3 (Kilchert *et al.*, 2016). Transcripts termed as CUTs were originally identified as ones that accumulate in *rrp6*Δ deletion mutant (Xu *et al.*, 2009) and comparison of the transcriptome of this mutant with that of Dis3 catalytic mutants revealed some unique and some specific roles of the two catalytic subunits (Gudipati *et al.*, 2012); however, the transcriptome of the Rrp6 catalytic mutant was studied only with *Schizosaccharomyces pombe* cells (Mukherjee *et al.*, 2016). The study with *S. pombe* revealed that some RNA targets of Rrp6 depended mainly on its structural role, such as RNAs of early meiotic and iron metabolism genes (Mukherjee *et al.*, 2016). Noncatalytic roles of Rrp6 have not yet been explored transcriptome-wide in *S. cerevisiae* but could be an interesting subject to study, especially with relation to unique cellular states, such as meiosis or conditions of heat shock.

The cell wall structure is absent from mammalian cells; however, protein glycosylation is conserved and essential for viability from yeast to human. The importance of protein glycosylation is underscored by the congenital disorders of glycosylation syndrome, which encompasses multisystemic diseases in children that result from defects in various steps along glycan modification pathways (Chang *et al.*, 2018). While final sugar composition and branching differs between yeast and human, the earliest steps in the glycosylation pathway, precursor synthesis and initial N-glycosylation reactions, are highly conserved (Lehle *et al.*, 2006). Our work in yeast clearly shows that one of the molecular consequences of RNA exosome inactivation is impairment of protein glycosylation at these early steps. Given the high conservation of both the RNA exosome complex and the glycosylation pathway, as well as the association of both with human diseases, this study opens the possibility for future investigation with human cells.

SUMMARY

RNA exosome activity, accomplished through Dis3 exonuclease activity and a noncatalytic function of Rrp6, is necessary for maintaining cell wall stability in yeast *Saccharomyces cerevisiae*.

A defect in protein glycosylation is a major reason for cell wall instability of RNA exosome mutants.

Genes encoding proteins involved in the early steps of protein glycosylation are dysregulated in RNA exosome mutants through mechanisms that involve increased accumulation of noncoding RNAs at high temperature.

MATERIALS AND METHODS

Request a protocol through *Bio-protocol*.

Strains, media, plasmids, and strain construction

Yeast strains and primers used in this study are listed in Supplemental Tables S1 and S2, respectively. Experiments were performed with the BMA41 (W303-derived) strain background, unless noted otherwise. Yeast strains were grown in YPD (containing per liter 20 g peptone, 10 g yeast extract, 20 g glucose, 0.1 g adenine) or YNB medium (containing per liter 6.7 g yeast nitrogen base without amino acids, 2 g drop-out mix as in Musladin *et al.*, 2014, 20 g glucose) supplemented with the required amino acids and uracil (80 mg/l each). Plasmid YEp352-*PSA1* is a high copy vector that carries the *PSA1/MPG1* gene (Janik *et al.*, 2003).

Psa1 was tagged at its genomic locus with a C-terminal 9xMyc tag. The tagging cassette was PCR amplified from plasmid pYM20 (Janke *et al.*, 2004) using the primer pair PSA1Ctag_fwd/ PSA1Ctag_rev and transformed into BMA41 wild-type and *rrp6*Δ strain by a standard lithium acetate procedure. Transformants were selected on Hygromycin B (0.3 mg/ml, Roche) plates and the presence of the tag was confirmed by Western blotting. The *RRP6* gene was deleted in BY4741 strain using a disruption cassette generated by PCR with primers RRP6-Kan1 and RRP6-Kan2 (Mosrin-Huaman *et al.*, 2009).

Phenotypic assays

Sensitivities to CR, CFW, caffeine, and SDS were tested by spotting assays. Exponential phase cultures were adjusted to an OD_{600} of 1 and four 10-fold serial dilutions of that sample were spotted onto plates supplemented with indicated amounts of each compound. Plates were incubated at 30°C or 37°C for 3 d and photographed using a UVIDOC HD6 camera (Uvitec, Cambridge).

Alkaline phosphatase activity assays

Activity of alkaline phosphatase released into the medium was measured as in Molina *et al.* (1998) with slight modifications. Supernatant (500 µl) from liquid culture was mixed with equal volume of 20 mM *p*-nitrophenylphosphate in Tris-HCl buffer, pH 8.8 and assayed for alkaline phosphatase activity. The reaction was performed at 30°C, stopped by the addition of 500 µl of 1 M NaOH, and absorbance of liberated *p*-nitrophenol was measured at 420 nm using a Helios Gamma spectrophotometer (Thermo Fisher). Enzyme activity was normalized to OD_{600} of the culture and the assay time in minutes and was expressed in arbitrary units: A_{420}*10,000/[OD_{600}*(*t*/min)*(V_{sample}/V_{total})]. Intracellular activity of alkaline phosphatase was measured exactly as described in Münsterkötter *et al.* (2000).

Fluorescence microscopy

Cells were stained with CFW stain (Sigma) and observed with an Olympus BX51 fluorescence microscope. The fluorescence from CFW was filtered with a DAPI filter.

RNA-seq data processing and computational analysis

Raw data from Wang *et al.* (2020) were downloaded via GEO (accession number GSE140504). Alignment and reads abundance estimation were conducted as described in the original publication. In short, Hisat2 was used to align reads against *S. cerevisiae* reference genome (taken from SGD, release R64-1-1); read abundance for mRNAs was estimated with HTseq-count (with the option –s reverse). Differential analysis between wt and *rrp6*Δ strain was conducted under the R environment using the DESeq2 package. Resulting log_2FC were used to construct heatmaps using the ggplot2 and complexHeatmap packages.

Analysis of the degree of glycosylation of periplasmic invertase

The secretory invertase was analyzed as described in Hashimoto *et al.* (1997). Briefly, invertase expression was induced by incubating midlogarithmic phase cells in medium that contains sucrose instead of glucose for 2 h at 30°C or 37°C. Cells were treated with zymolyase and the periplasmic fraction containing invertase was separated from spheroplasts by centrifugation. The periplasmic fraction was subjected to 7.5% SDS–PAGE, gel was bathed in 0.1 M sodium acetate, pH 5.1, containing 0.1 M sucrose at 37°C for 1 h to carry out the enzymatic reaction of invertase, and then washed with water, placed in 0.1% 2,3,5- triphenyltetrazolium chloride, 0.5 M NaOH, and boiled to detect red bands. After staining, gel was washed with 7.5% acetic acid.

RNA isolation and RT-qPCR analysis

Total RNA was extracted by the hot phenol method (Schmitt *et al.*, 1990) and column purified with DNase treatment using a NucleoSpin RNA kit (Macherey Nagel) according to manufacturer instructions. RNA was quantified with a Nanodrop spectrophotometer and 1 µg was used in a strand-specific reverse-transcription reaction with a ProtoScript First Strand cDNA Synthesis Kit (New England Biolabs) with 0.1 µM gene-specific oligonucleotides and supplemented with actinomycin D (Sigma) to final concentration 5 µg/ml to ensure strand specificity. Twofold diluted cDNA (1 µl) was then amplified in Roche LightCycler 480 with the Maxima SYBR Green qPCR Master Mix detection kit from Thermo Scientific as recommended by the supplier. The qPCR datasets were analyzed using the ΔΔCt method, and the results were normalized to *PMA1* mRNA RT-qPCR amplification, which was used as internal control. The level of a certain transcript for each sample was expressed relative to its abundance in wild-type cells at 30°C, which was set as 1. Amplifications were done in duplicate for each sample, and three independent RNA extractions were analyzed.

Western blot analysis

Total proteins were obtained as described in Kushnirov (2000), resolved on SDS 10% polyacrylamide electrophoresis gels, and analyzed by Western blotting according to standard procedures. Myc-tagged Psa1 was probed with anti-c-Myc (9E10; Santa Cruz Biotechnology) at 1:1000 dilution and Pgk1 with anti-PGK1 (22C5D8; Abcam) at 1:5000 dilution. In both cases, mouse IgG kappa-binding protein HRP (Santa Cruz Biotechnology) at 1:50,000 dilution was used to detect the primary antibody. Blots were developed using Biorad Clarity Western ECL substrates and visualized with a C-DiGit Blot scanner (LI-COR Biosciences). Band intensity was quantified with GelAnalyzer 19.1 software and the results were normalized to Pgk1. The level of protein was expressed relative to its abundance in wild-type cells at 30°C, which was set as 1. Three independent protein extractions were analyzed for each sample.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed similarly as described in Stuparevic *et al.* (2013). Forty milliliters of cells were fixed with 1% formaldehyde for 20 min. After glycine addition to stop the reaction, the cells were washed and lysed with glass beads to isolate chromatin. The cross-linked chromatin was sheared by sonication with a Vibra-Cell sonicator to reduce average fragment size to approximately 500 base pairs. Chromatin fractions of 400 µl were taken for each immunoprecipitation reaction and incubated with 4 µl of anti-RNA polymerase II antibodies (8WG16, sc-56767; Santa Cruz Biotechnology) at 4°C overnight. After incubation, 40 µl of protein G PLUS-agarose beads (sc-2002; Santa Cruz Biotechnology) were added and incubated for 2 h at 4°C. The beads were then washed extensively, and the chromatin was eluted. Eluted supernatants (output) and the input controls were hydrolyzed with Pronase (0.8 mg/ml final concentration; Sigma) for 2 h at 42°C, followed by 7 h incubation at 65°C to reverse cross-linked DNA complexes. DNA was extracted using the Macherey Nagel Nucleospin Gel & PCR Cleanup Kit. The immunoprecipitated DNAs (output) were quantified by qPCR in Roche LightCycler 480 with the Maxima SYBR Green qPCR Master Mix detection kit from Thermo Scientific as recommended by the supplier. Amplifications were done in triplicate for each sample. Immunoprecipitated samples (output) were normalized to input.

ACKNOWLEDGMENTS

This work was supported by Croatian Science Foundation Grant no. UIP-2017-05-4411 to I.S., funding from la Ligue Contre le Cancer (comités du Loiret et de l'Indre) to A.R.R., and funding from the Croatian Academy of Sciences and Arts to A.N. V.B. is a recipient of a PhD fellowship from the Région Centre-Val de Loire and A.N.'s research stay at the CBM, CNRS in France was funded by a Scholarship of the French Government. We are grateful to Michael Primig and Vladimir Mrša for helpful discussions. We are also grateful to Sandi Orlić for help with fluorescence microscopy and to Grazyna Palamarczyk for the gift of YEp352-*MPG1* plasmid.

REFERENCES

Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D (1999a). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J 18, 5399–5410.

Allmang C, Mitchell P, Petfalski E, Tollervey D (2000). Degradation of ribosomal RNA precursors by the exosome. Nucleic Acids Res 28, 1684–1691.

Allmang C, Petfalski E, Podtelejnikov A, Mann M, Tollervey D, Mitchell P (1999b). The yeast exosome and human PM-Scl are related complexes of 3′ → 5′ exonucleases. Genes Dev 13, 2148–2158.

Amorim J De, Slavotinek A, Fasken MB, Corbett AH, Morton DJ, Corbett AH (2020). Modeling pathogenic variants in the RNA exosome. RNA Dis 7, 1–9.

Belcarz A, Ginalska G, Lobarzewski J, Penel C (2002). The novel non-glycosylated invertase from *Candida utilis* (the properties and the conditions of production and purification). Biochim Biophys Acta - Protein Struct Mol Enzymol 1594, 40–53.

Benton BK, Plump SD, Roos J, Lennarz WJ, Cross FR (1996). Over-expression of *S. cerevisiae* G1 cyclins restores the viability of alg1 N-glycosylation mutants. Curr Genet 29, 106–113.

Bresson S, Tuck A, Staneva D, Tollervey D (2017). Nuclear RNA decay pathways aid rapid remodeling of gene expression in yeast. Mol Cell 65, 787–800.e5.

Briggs MW, Burkard KTD, Butler JS (1998). Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5. 8 S rRNA 3. J Biol Chem 273, 13255–13263.

Callahan KP, Butler JS (2008). Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p. Nucleic Acids Res 36, 6645–6655.

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

Catala M, Aksouh L, Elela SA (2012). RNA-dependent regulation of the cell wall stress response. Nucleic Acids Res 40, 7507–7517.

Chang IJ, He M, Lam CT (2018). Congenital disorders of glycosylation. Ann Transl Med 6, 477.

Chlebowski A, Lubas M, Jensen TH, Dziembowski A (2013). RNA decay machines: The exosome. Biochim Biophys Acta - Gene Regul Mech 1829, 552–560.

Davidson L, Francis L, Cordiner RA, Eaton JD, Estell C, Macias S, Cáceres JF, West S (2019). Rapid depletion of DIS3, EXOSC10, or XRN2 reveals the immediate impact of exoribonucleolysis on nuclear RNA metabolism and transcriptional control. Cell Rep 26, 2779–2791.e5.

Delan-Forino C, Spanos C, Rappsilber J, Tollervey D (2020). Substrate specificity of the TRAMP nuclear surveillance complexes. Nat Commun 11, 3122–3137.
Drązkowska K, Tomecki R, Stoduś K, Kowalska K, Czarnocki-Cieciura M, Dziembowski A (2013). The RNA exosome complex central channel controls both exonuclease and endonuclease Dis3 activities in vivo and in vitro. Nucleic Acids Res 41, 3845–3858.
Dziembowski A, Lorentzen E, Conti E, Séraphin B (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat Struct Mol Biol 14, 15–22.
Esmon PC, Esmon BE, Schauer IE, Taylor A, Schekman R (1987). Structure, assembly, secretion of octameric invertase. J Biol Chem 262, 4387–4394.
Fasken MB, Morton DJ, Kuiper EG, Jones SK, Leung SW, Corbett AH (2020). The RNA exosome and human disease. Methods Mol Biol 2062, 3–33.
Feigenbutz M, Garland W, Turner M, Mitchell P (2013). The exosome cofactor Rrp47 is critical for the stability and normal expression of its associated exoribonuclease Rrp6 in *Saccharomyces cerevisiae*. PLoS One 8, 1–17.
Gudipati RK, Xu Z, Lebreton A, Séraphin B, Steinmetz LM, Jacquier A, Libri D (2012). Extensive degradation of RNA precursors by the exosome in wild-type cells. Mol Cell 48, 409–421.
Hainer SJ, Pruneski JA, Mitchell RD, Monteverde RM, Martens JA (2011). Intergenic transcription causes repression by directing nucleosome assembly. Genes Dev 25, 29–40.
Hashimoto H, Sakakibara A, Yamasaki M, Yoda K (1997). *Saccharomyces cerevisiae* VIG9 encodes GDP-mannose pyrophosphorylase, which is essential for protein glycosylation. J Biol Chem 272, 16308–16314.
Hilleren P, McCarthy T, Rosbash M, Parker R, Jensen TH (2001). Quality control of mRNA 3′-end processing is linked to the nuclear exosome. Nature 413, 538–542.
Janik A, Juchimiuk M, Kruszewska J, Orowski J, Pasikowska M, Palamarczyk G (2012). Impact of yeast glycosylation pathway on cell integrity and morphology. Glycosylation 11, 259–272.
Janik A, Sosnowska M, Kruszewska J, Krotkiewski H, Lehle L, Palamarczyk G (2003). Overexpression of GDP-mannose pyrophosphorylase in *Saccharomyces cerevisiae* corrects defects in dolichol-linked saccharide formation and protein glycosylation. Biochim Biophys Acta - Gen Subj 1621, 22–30.
Janke C, Magiera MM, Rathfelder N, Taxis C, Reber S, Maekawa H, Moreno-Borchart A, Doenges G, Schwob E, Schiebel E, *et al.* (2004). A versatile toolbox for PCR-based tagging of yeast genes: New fluorescent proteins, more markers and promoter substitution cassettes. Yeast 21, 947–962.
Kilchert C, Wittmann S, Vasiljeva L (2016). The regulation and functions of the nuclear RNA exosome complex. Nat Rev Mol Cell Biol 17, 227–239.
Klauer AA, Van Hoof A (2013). Genetic interactions suggest multiple distinct roles of the arch and core helicase domains of Mtr4 in Rrp6 and exosome function. Nucleic Acids Res 41, 533–541.
Klis FM, Mol P, Hellingwerf K, Brul S (2002). Dynamics of cell wall structure in *Saccharomyces cerevisiae*. FEMS Microbiol Rev 26, 239–256.
Kopecká M, Gabriel M (1992). The influence of Congo red on the cell wall and (1 → 3)-β-d-glucan microfibril biogenesis in *Saccharomyces cerevisiae*. Arch Microbiol 158, 115–126.
Kuranda K, Leberre V, Sokol S, Palamarczyk G, François J (2006). Investigating the caffeine effects in the yeast *Saccharomyces cerevisiae* brings new insights into the connection between TOR, PKC and Ras/cAMP signalling pathways. Mol Microbiol 61, 1147–1166.
Kushnirov VV (2000). Rapid and reliable protein extraction from yeast. Yeast 16, 857–860.
LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell 121, 713–724.
Lardenois A, Liu Y, Walther T, Chalmel F, Evrard B, Granovskaia M, Chu A, Davis RW, Steinmetz LM, Primig M (2011). Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. Proc Natl Acad Sci 108, 1058–1063.
Lebreton A, Tomecki R, Dziembowski A, Séraphin B (2008). Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature 456, 993–996.
Lehle L, Strahl S, Tanner W (2006). Protein glycosylation, conserved from yeast to man: A model organism helps elucidate congenital human diseases. Angew Chemie - Int Ed 45, 6802–6818.
Lingaraju M, Schuller JM, Falk S, Gerlach P, Bonneau F, Basquin J, Benda C, Conti E (2020). To process or to decay: a mechanistic view of the nuclear RNA exosome. Cold Spring Harb Symp Quant Biol LXXXIV, 040295.
Makino DL, Schuch B, Stegmann E, Baumgärtner M, Basquin C, Conti E (2015). RNA degradation paths in a 12-subunit nuclear exosome complex. Nature 524, 54–58.
Milbury KL, Paul B, Lari A, Fowler C, Montpetit B, Stirling PC (2019). Exonuclease domain mutants of yeast DIS3 display genome instability. Nucleus 10, 1–12.
Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D (1997). The exosome: A conserved eukaryotic RNA processing complex containing multiple 3′→5′ exoribonucleases. Cell 91, 457–466.
Molina M, Martín H, Sánchez M, Nombela C (1998). MAP kinase-mediated signal transduction pathways. Methods Microbiol 26, 375–393.
Moreau K, Le Dantec A, Mosrin-Huaman C, Bigot Y, Piégu B, Rahmouni AR (2019). Perturbation of mRNP biogenesis reveals a dynamic landscape of the Rrp6-dependent surveillance machinery trafficking along the yeast genome. RNA Biol 16, 879–889.
Mosrin-Huaman C, Honorine R, Rahmouni AR (2009). Expression of bacterial Rho factor in yeast identifies new factors involved in the functional interplay between transcription and mRNP biogenesis. Mol Cell Biol 29, 4033–4044.
Mukherjee K, Gardin J, Futcher B, Leatherwood J (2016). Relative contributions of the structural and catalytic roles of Rrp6 in exosomal degradation of individual mRNAs. RNA 22, 1311–1319.
Münsterkötter M, Barbaric S, Hörz W (2000). Transcriptional regulation of the yeast PHO8 promoter in comparison to the coregulated PHO5 promoter. J Biol Chem 275, 22678–22685.
Musladin S, Krietenstein N, Korber P, Barbaric S (2014). The RSC chromatin remodeling complex has a crucial role in the complete remodeler set for yeast PHO5 promoter opening. Nucleic Acids Res 42, 4270–4282.
Novačić A, Vučenović I, Primig M, Stuparević I (2020). Non-coding RNAs as cell wall regulators in *Saccharomyces cerevisiae*. Crit Rev Microbiol 0, 1–11.
Orlean P (2012). Architecture and biosynthesis of the *Saccharomyces cerevisiae* cell wall. Genetics 192, 775–818.
Phillips S, Butler JS (2003). Contribution of domain structure to the RNA 3′ end processing and degradation functions of the nuclear exosome subunit Rrp6p. RNA 9, 1098–1107.
Popolo L, Gualtieri T, Ragni E (2001). The yeast cell-wall salvage pathway. Med Mycol 39, 111–121.
Roncero C, Duran A (1985). Effect of Calcofluor white and Congo red on fungal cell wall morphogenesis: in vivo activation of chitin polymerization. J Bacteriol 163, 1180–1185.
Schilders G, Raijmakers R, Raats JMH, Pruijn GJM (2005). MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation. Nucleic Acids Res 33, 6795–6804.
Schmidt K, Xu Z, Mathews DH, Butler JS (2012). Air proteins control differential TRAMP substrate specificity for nuclear RNA surveillance. RNA 18, 1934–1945.
Schmitt ME, Brown TA, Trumpower BL (1990). A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res 18, 3091–3092.
Schneider C, Kudla G, Wlotzka W, Tuck A, Tollervey D (2012). Transcriptome-wide analysis of exosome targets. Mol Cell 48, 422–433.
Schroeder L, Ikui AE (2019). Tryptophan confers resistance to SDS-associated cell membrane stress in *Saccharomyces cerevisiae*. PLoS One 14, 1–16.
Stuparevic I, Mosrin-Huaman C, Hervouet-Coste N, Remenaric M, Rahmouni AR (2013). Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP) in yeast. J Biol Chem 288, 31816–31829.
Tomlin GC, Hamilton GE, Gardner DCJ, Walmsley RM, Stateva LI, Oliver SG (2000). Suppression of sorbitol dependence in a strain bearing a mutation in the SRB1/PSA1/VIG9 gene encoding GDP-mannose pyrophosphorylase by PDE2 overexpression suggests a role for the Ras/cAMP signal-transduction pathway in the control of yeast cell-wall biogenesis. Microbiology 146, 2133–2146.
Trachtulcová P, Frýdlová I, Janatová I, Dorosh A, Hašek J (2003). The W303 genetic background affects the isw2Δ mutant phenotype in *Saccharomyces cerevisiae*. Folia Microbiol (Praha) 48, 745–753.
Villa T, Barucco M, Martin-Niclos MJ, Jacquier A, Libri D (2020). Degradation of non-coding RNAs promotes recycling of termination factors at sites of transcription. Cell Rep 32, 107942–107957.

Wang C, Liu Y, DeMario SM, Mandric I, Gonzalez-Figueroa C, Chanfreau GF (2020). Rrp6 Moonlights in an RNA exosome-independent manner to promote cell survival and gene expression during stress. Cell Rep 31, 107754.
Warit S, Zhang N, Short A, Walmsley RM, Oliver SG, Stateva LI (2000). Glycosylation deficiency phenotypes resulting from depletion of GDP-mannose pyrophosphorylase in two yeast species. Mol Microbiol 36, 1156–1166.
Wasmuth EV, Januszyk K, Lima CD (2014). Structure of an Rrp6-RNA exosome complex bound to poly(A) RNA. Nature 511, 435–439.
Wasmuth EV, Lima CD (2012). Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. Mol Cell 48, 133–144.
Wasmuth EV, Lima CD (2017). The Rrp6 C-terminal domain binds RNA and activates the nuclear RNA exosome. Nucleic Acids Res 45, 846–860.
Wasmuth EV, Zinder JC, Zattas D, Das M, Lima CD (2017). Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that mpp6 stimulates RNA decay and recruits the Mtr4 helicase. Elife 6, 1–24.
Van Dijk EL, Schilders G, Pruijn GJM (2007). Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways. RNA 13, 1027–1035.
Van Werven FJ, Neuert G, Hendrick N, Lardenois A, Buratowski S, Van Oudenaarden A, Primig M, Amon A (2012). Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell 150, 1170–1181.
Winston F, Martens JA, Wu P-YJ (2005). Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. Genes Dev 19, 2695–2704.
Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, *et al*. (2005). Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121, 725–737.
Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W, Steinmetz LM (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature 457, 1033–1037.
Yu Y, Yarrington RM, Chuong EB, Elde NC, Stillman DJ (2016). Disruption of promoter memory by synthesis of a long noncoding RNA. Proc Natl Acad Sci USA 113, 9575–9580.
Zhang N, Gardner DCJ, Oliver SG, Stateva LI (1999). Down-regulation of the expression of PKC1 and SRB1/PSA1/VIG9, two genes involved in cell wall integrity in *Saccharomyces cerevisiae*, causes flocculation. Microbiology 145, 309–316.
Zinder JC, Lima CD (2017). Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors. Genes Dev 31, 88–100.

Table S1. *S. cerevisiae* strains

Strain ID	Genotype	Source
BMA41 wild type	*MATa ade2-1 ura3-1 leu2-3,112 his3-11,15 trp1Δ can1-100*	(Baudin-Baillieu *et al.*, 1997)
BMA41 *rrp6Δ*	BMA41 with *rrp6::KanMX4*	(Mosrin-Huaman *et al.*, 2009)
BMA41 Psa1-Myc	BMA41 with *PSA1-Myc::hph*	This work
BMA41 *rrp6Δ* Psa1-Myc	BMA41 *rrp6* with *PSA1-Myc::hph*	This work
BMA41 Rrp6-Y361A	BMA41 with *rrp6Y361A*	(Stuparevic *et al.*, 2013)
BMA41 *rrp47Δ*	BMA41 with *rrp47::KanMX4*	(Stuparevic *et al.*, 2013)
BMA41 *trf4Δ*	BMA41 with *trf4::KanMX4*	(Stuparevic *et al.*, 2013)
BMA41 *trf5Δ*	BMA41 with *trf5::KanMX4*	(Stuparevic *et al.*, 2013)
BMA41 *mpp6Δ*	BMA41 with *mpp6::KanMX4*	(Stuparevic *et al.*, 2013)
BMA41 *air1Δ*	BMA41 with *air1::KanMX4*	(Stuparevic *et al.*, 2013)
BMA41 *air2Δ*	BMA41 with *air2::KanMX4*	(Stuparevic *et al.*, 2013)
BMA41 *air1Δ air2Δ*	*MATa ade2-1 ura3-1 leu2-3,112 his3-11,15 trp1Δ can1-100 air1::HIS3 air2::KanMX4*	(Mosrin-Huaman *et al.*, 2009)
DIS3	*MATa ade2-1 ura3-1 leu2-3,112 his3-11,15 trp1-1 can1-100 dis3::KanMX4 [pBS3269-DIS3, LEU2]*	(Stuparevic *et al.*, 2013)
dis3 endo^{-}	*MATa ade2-1 ura3-1 leu2-3,112 his3-11,15 trp1-1 can1-100 dis3::KanMX4 [pBS3278-dis3D171N, LEU2]*	(Stuparevic *et al.*, 2013)
dis3 exo^{-}	*MATa ade2-1 ura3-1 leu2-3,112 his3-11,15 trp1-1 can1-100 dis3::KanMX4 [pBS3270-dis3D551N, LEU2]*	(Stuparevic *et al.*, 2013)
BY4741 wild type	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	(Brachmann *et al.*, 1998)
BY4741 *rrp6Δ*	BY4741 with *rrp6::KanMX4*	This work
JHY222 wild type	*MATa/MATα HAP1/HAP1 MKT1(D30G)/MKT1(D30G) RME1(INS 308A)/RME1(INS 308A) TAO3(E1493Q)/TAO3(E1493Q)*	(Lardenois *et al.*, 2011)
JHY222 *rrp6Δ/rrp6Δ*	JHY222 with *rrp6::KanMX4/rrp6::KanMX4*	(Lardenois *et al.*, 2011)

Table S2. Primers

Primer ID	Sequence (5'→3')	Description
RRP6-Kan1	GATAGACGAAATAGGAACAACAAACAGCTTATAA GCACCCAATAAGTGCGTT CCCGGCCAGCGACATGGAGGCCCAG	Deletion of *RRP6*
RRP6-Kan2	GCCCTTGGTCCATTACTATCGCTAGATGATGGGT CGAATCTCCTTTTCCGAATCG ACAGCAGTATAGCGACCAGGCT	Deletion of *RRP6*
PSA1Ctag_fwd	TACCTCATAAGTCTATCTCCGATAATGTTCCAAAG GAAGCTATTATTATGCGTACGCTGCAGGTCGAC	C-terminal tagging of Psa1
PSA1Ctag_rev	ACAGATGAGTGATATAGATTATATTCATTACAGTT CGTTTTCCTAACTCAATCGATGAATTCGAGCTCG	C-terminal tagging of Psa1
PSA1_fwd	GATTGGCCCAGACGTGGTTA	RT-qPCR
PSA1_rev	ACTTCAACGTCGTCACCCAA	RT-qPCR
CUT488_fwd	GCTGGTGGTTTCTCGCTTTG	RT-qPCR
CUT488_rev	TGTGAAACTTCTAACAATGACGTG	RT-qPCR
PMA1_fwd	GGCTTCCATAACGAATTGAATTGGACCG	RT-qPCR
PMA1_rev	CAATCTAATCACGGTGTCGACGACGAAGAC	RT-qPCR
DPM1_fwd	TGGCCAGACCTTTGACCATC	RT-qPCR
DPM1_rev	CCTTGCGAGTTGATGTCCCT	RT-qPCR
ALG7_fwd	TTCAAATTGGTCCCCTGCCC	RT-qPCR
ALG7_rev	TGCTCTTTGGCGGTTCTTCT	RT-qPCR
PSA1prom_fwd	TGTTCTCACCTCTTTGACTTTGA	ChIP qPCR
PSA1prom_rev	GTTGTAGCTTGTCTTGTGCTGA	ChIP qPCR
TAF10prom_fwd	AGCTGCAGATTCAGCATTCA	ChIP qPCR
TAF10prom_rev	TCCGCATCGTAATCTTCCTCA	ChIP qPCR

REFERENCES

Baudin-Baillieu, A, Tollervey, D, Cullin, C, and Lacroute, F (1997). Functional analysis of Rrp7p, an essential yeast protein involved in pre-rRNA processing and ribosome assembly. Mol Cell Biol 17, 5023–5032.

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

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

Mosrin-Huaman, C, Honorine, R, and Rahmouni, AR (2009). Expression of bacterial Rho factor in yeast identifies new factors involved in the functional interplay between transcription and mRNP biogenesis. Mol Cell Biol 29, 4033–4044.

Stuparevic, I, Mosrin-Huaman, C, Hervouet-Coste, N, Remenaric, M, and Rahmouni, AR (2013). Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP) in yeast. J Biol Chem 288, 31816–31829.

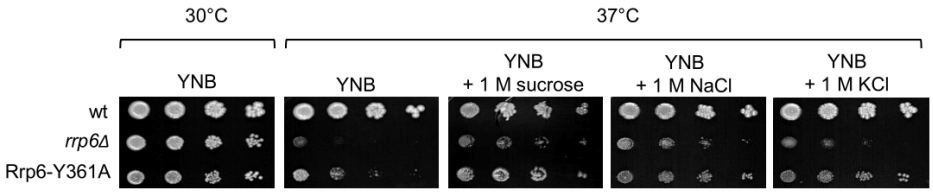


Figure S1. Osmotic stabilization of *rrp6Δ* cells at high temperature happens regardless of osmotic stabilizer used. 10-fold serial dilutions of BMA41 wild type (wt) and isogenic mutant strains cells were spotted on plates and were photographed after 3 days at indicated temperature.

Figure S1

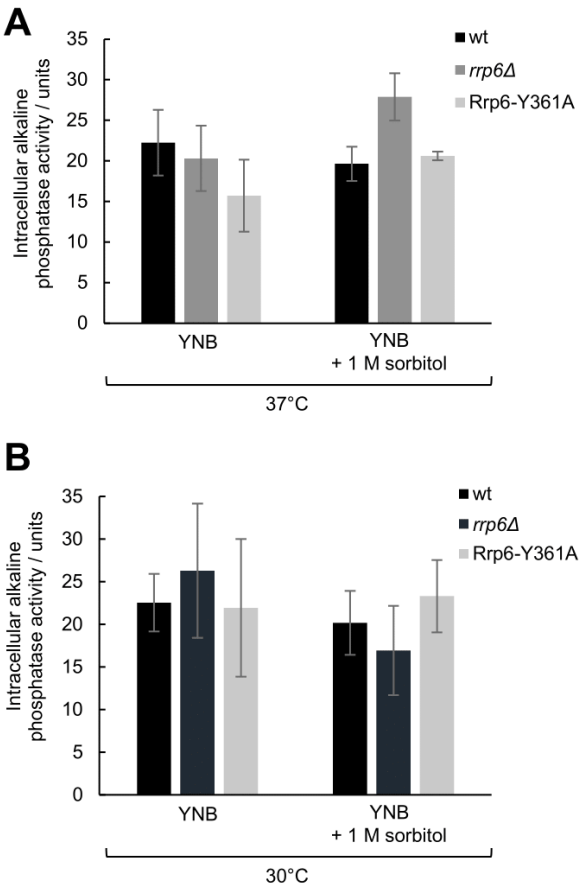


Figure S2. Alkaline phosphatase is not differentially expressed in *rrp6Δ* or Rrp6-Y361A mutant. BMA41 wild type (wt) and isogenic mutant strains were grown in YNB medium for 3 days at 37°C (A) or 30°C (B) and intracellular alkaline phosphatase activity was measured. Measurements were performed in duplicate, and reported values represent the means and standard deviations of three independent experiments (n=3).

Figure S2

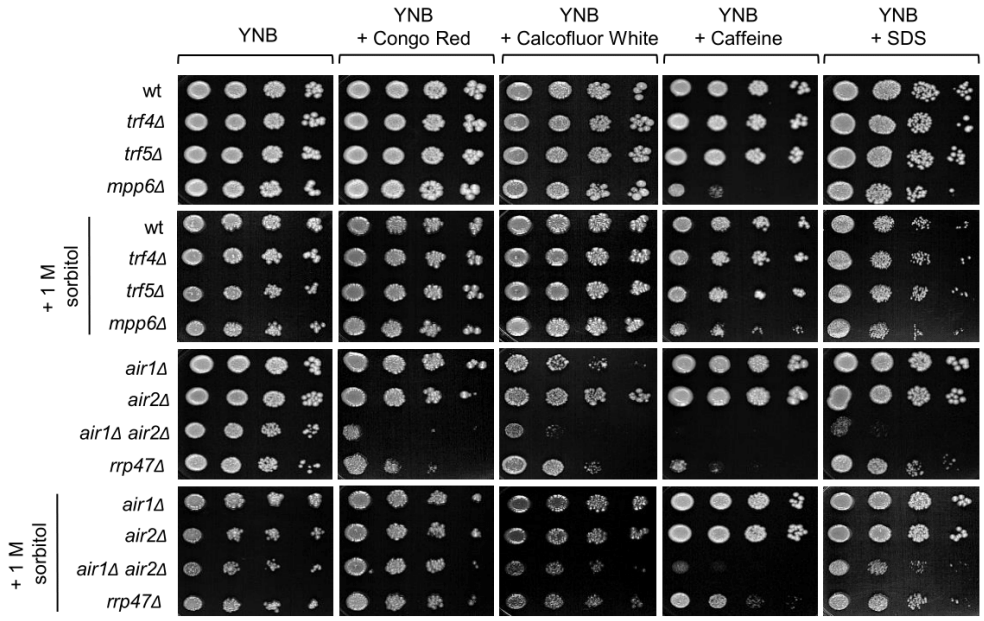


Figure S3. Temperature-sensitive mutants in RNA exosome cofactors are hypersensitive to cell wall stressors. Strains are described in Figure 2. 10-fold serial dilutions of cells were spotted on plates and were photographed after 3 days at 30°C. Concentrations of compounds used: Congo Red 10 µg/ml, Calcofluor White 20 µg/ml, Caffeine 6 mM, SDS 0,0075%.

Figure S3

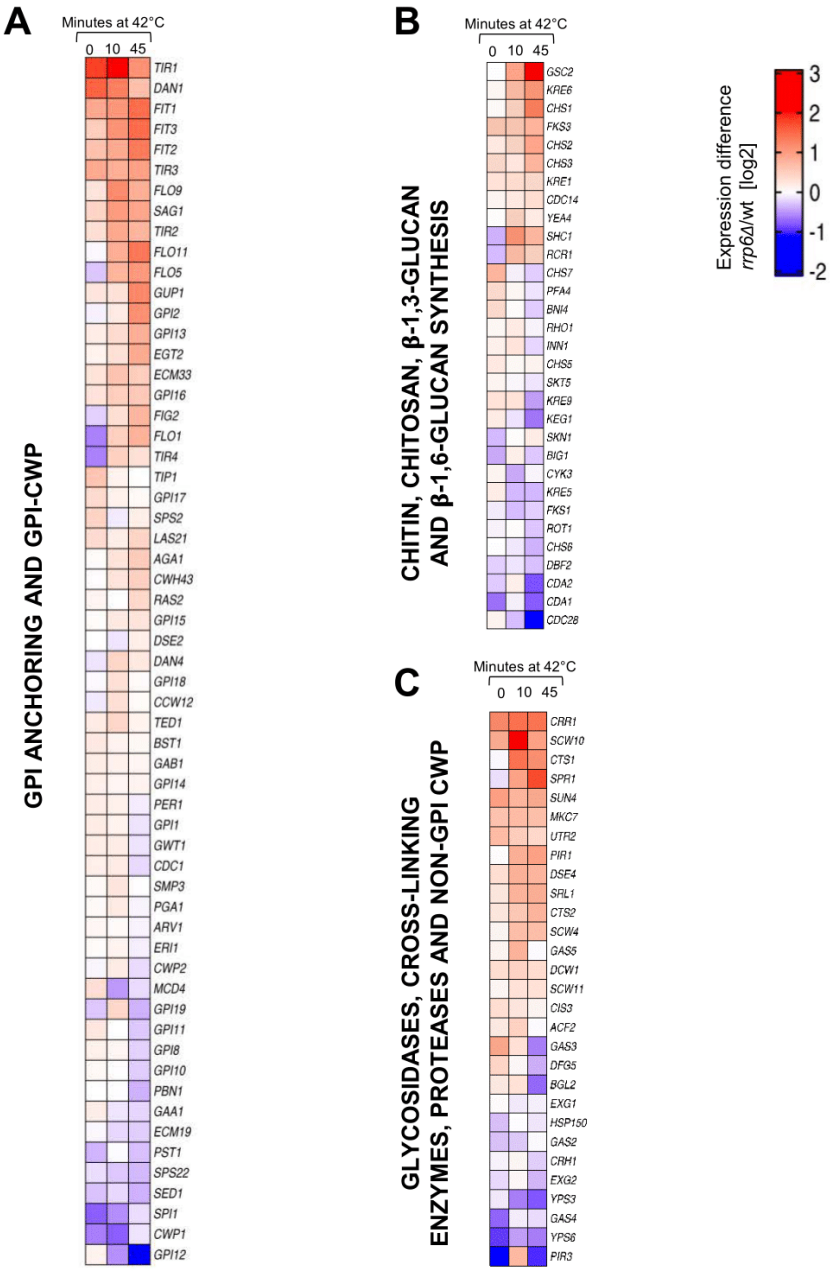


Figure S4. Expression profiles of mRNAs encoding cell wall-related proteins in cells lacking Rrp6 during heat shock. RNA-seq heat map with data from (Wang et al. 2020) showing the expression differences of mRNAs as *rrp6Δ*/wt mRNA ratio on a log2 scale, for genes subdivided into cell wall-related gene categories adapted from (Orlean 2012): (A) GPI anchoring and GPI-CWP (cell wall proteins), (B) Chitin, chitosan, **β**-1,3 glucan and **β**-1,6 glucan synthesis and (C) Glycosidases, cross-linking enzymes, proteases and non-GPI CWP.

Figure S4

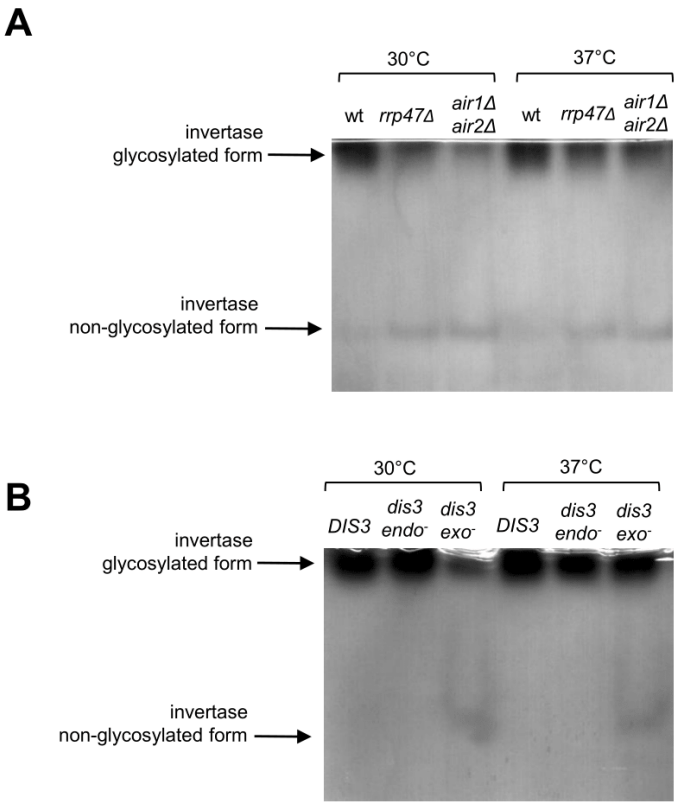


Figure S5. Protein glycosylation is affected in RNA exosome mutants. Strains are described in Figure 2. Activity staining of invertase from periplasmic extracts of *rrp47Δ* and *air1Δair2Δ* cells (A), as well as *dis3 exo-* cells (B) contain an additional non-glycosylated form of periplasmic invertase.

Figure S5

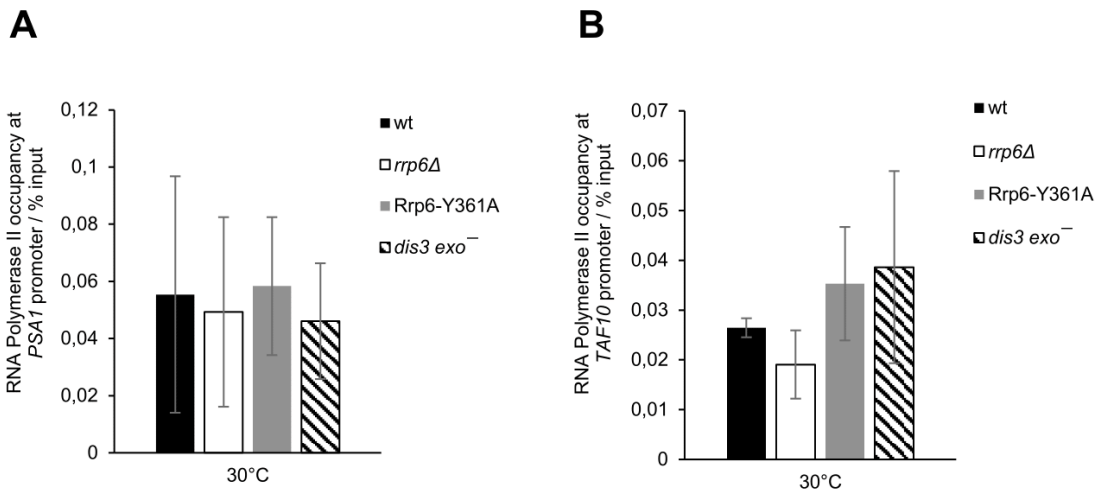


Figure S6. There is no difference in recruitment of RNA Polymerase II to *PSA1* or *TAF10* gene promoters in RNA exosome mutants compared to wild type cells at physiological temperature. Strains are described in Figures 1 and 2. Quantification was performed by ChIP of RNA polymerase II using specific antibodies 8WG16. Immunoprecipitated samples (output) were normalized to input following quantification by qPCR. Reported values represent the means and range of two independent experiments (n=2).

Figure S6

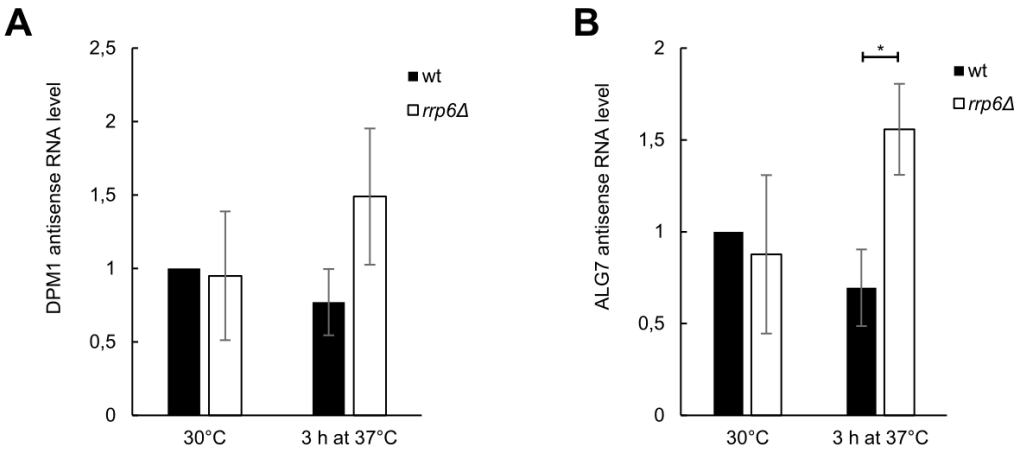


Figure S7. DPM1 and ALG7 antisense RNAs accumulate in cells lacking Rrp6 at high temperature. RT-qPCR values for DPM1 (A) and ALG7 (B) antisense RNAs for BMA41 wild type (wt) and isogenic *rrp6Δ* mutant strain are normalized to PMA1 mRNA and expressed relative to transcript abundance in wild type cells at 30°C, which is set as 1. Reported values represent the means and standard deviations of three independent experiments (n=3). Indicated differences show the significant differences using an unpaired student T-test. One (*) asterisk denotes a p-value lower or equal to 0.05.

Figure S7

2.3 Ongoing work

2.3.1 Introduction

For nearly a decade, we have studied the degradation of aberrant mRNP in yeast by using the bacterial Rho factor. Recently, we added HTS to our approach with the Rho model to explore further the RNA degradation mechanism. We have precisely identified the transcripts affected by Rho and discovered that some proteins involved in ncRNA degradation are recruited to loci coding Rho-affected transcripts. In this work, we aim to expand our approach to human cells.

In yeast, as in humans, the study of the RNA degradation mechanism is arduous due to the small number of aberrant mRNPs generated in physiological conditions. Therefore, the study of the RNA degradation mechanism is performed in strains lacking one or more proteins of the exosome or its cofactors (NEXT, PAXT) [185, 194, 238, 239]. Here, we propose an alternative using the bacterial Rho factor as a tool to induce the formation of aberrant mRNPs. The larger number of aberrant mRNP is therefore easier to track, even without typical inhibition of the exosome or its cofactors. We explored the effect of Rho on the transcription, processing and degradation of RNAs in HeLa cells. We were initially interested in the expression of some selected mRNA coding genes and mRNPs fate after the induction of Rho. We also focused on the emergence of a nuclear structure similar to the one discovered in [240]. We then broadened our analysis by using RNA-seq to identify transcripts affected by Rho genome-wide and their fate. In addition, we aimed to determine which transcripts are retained in the nucleus and the mechanisms determining this retention.

2.3.2 Material and methods

Generation of mammalian cell lines expressing Rho: The protocols used to produce the transfected plasmid and the stable Rho expressing HeLa cell lines are adapted from [218]. Briefly, the open-reading frames (ORF) of Rho-NLS were amplified by PCR from the yeast plasmid and modified to add HindIII and XhoI restriction sites. Modified ORF was then inserted in the polylinker of the pIND mammalian expression vector (Invitrogen). The protocol for the conception for the yeast plasmid can be found here [25]

HeLa cells harboring the regulatory ecdysone-inducible mammalian expression system within the vector pVgRXR (Invitrogen) were transfected using a plasmid/PEI mix. Stable transfectants were then selected by dilution cloning based on growth in the presence of G418 at 800 µg/ml (resistance gene in pIND) and zeocin at 400 µg/ml (resistance gene in pVgRXR). The growing clones were first verified by PCR-amplicon sequencing then confirmed by Western blot analyses of Rho expression induced after addition of Ponasterone A (Sigma Aldrich ref: P3490) into the medium at 5 µg/ml. Among many positive clones, appropriate cell lines exhibiting the best repression and inducibility of Rho were chosen.

RNA extraction: RNA extraction was performed using the NucleoSpin RNA-plus kit from Macherey-Nagel following the standard protocol from the manufacturer.

Buffers : SSPE 20X (1L) : 175.3g NaCl, 27.6g NaH2PO4.H2O, 7.4g EDTA 800mL, adjusting pH to 7.4 with NaOH then completing to 1L. Hybridization buffer (400 µL): 120µL SSPE20X, 40µL Denhard 50X, 10µL tRNA 20µg/µL, 230µL H2O

FISH and immunofluorescence: Cells were fixated and permeabilized by incubation at -20°C for 7 min in 90% methanol. RNA was detected using Cy3 labeled DNA oligonucleotides designed against polyA tails. Hybridization was performed by incubating the cells and the denatured probes in hybridization buffer for 1h at room temperature in a dark room. Each sample was then washed 3 times with SSPE 6X + 0.05% Tween 20 for 5min. DAPI coloration was performed following standard protocol from the manufacturer. For immunofluorescence detection of proteins, methanol fixed cells were stained using primary antibodies against RRP6, RRP40 or SC-35. Primary antibodies were targeted by a secondary antibody labeled with fluorescein.

qPCR: qPCR was performed following a protocol in [216].

RNA-seq analysis: The RNA-seq was analyzed via a typical workflow (alignment, read counting and differential analysis). Quality control of reads and correlation between replicates were performed on the initial fastq sequences and on post-alignment files respectively. The alignment was performed using HISAT2 on the human GRCh38 genome from NCBI. Counts on total length genes were performed to check the overall expression level regardless of splicing. Additional counts on all transcripts (intron excluded) were also produced to study eventual splicing defects that

could arise from Rho induction. The counts table was produced using FeatureCounts with all optional files provided to optimize the handling of exon junctions by the program. All further analyses were conducted under R environment. Specifically, EdgeR was used to measure differential expression between conditions using each replicate. Any expression variation with an insufficient p-value (>0.05) was discarded. Only genes whose transcripts were detected in both the cytosol and the nucleus were kept. Genes that did not show any expression variation for the two conditions (NI/12h and 12h/24h) were discarded.

2.3.3 Results

2.3.3.1 Expression of Rho in HeLa cells leads to different expression variations depending on induction time

Initial tests were conducted using HeLa cells transfected with a plasmid containing the Rho coding sequence coupled with a nuclear localization signal. Similar to its effects in yeast, we expected that Rho induction would lead to drastic augmentation of aberrant mRNP levels in the nucleus.

Measurement of the expression of four housekeeping genes after 12 and 24h of Rho induction showed a decrease in the first 12h followed by a large increase during the next 12h. Interestingly, the final level of mRNA after 24h of Rho induction exceeded the basal level measured at T=0 (NI).

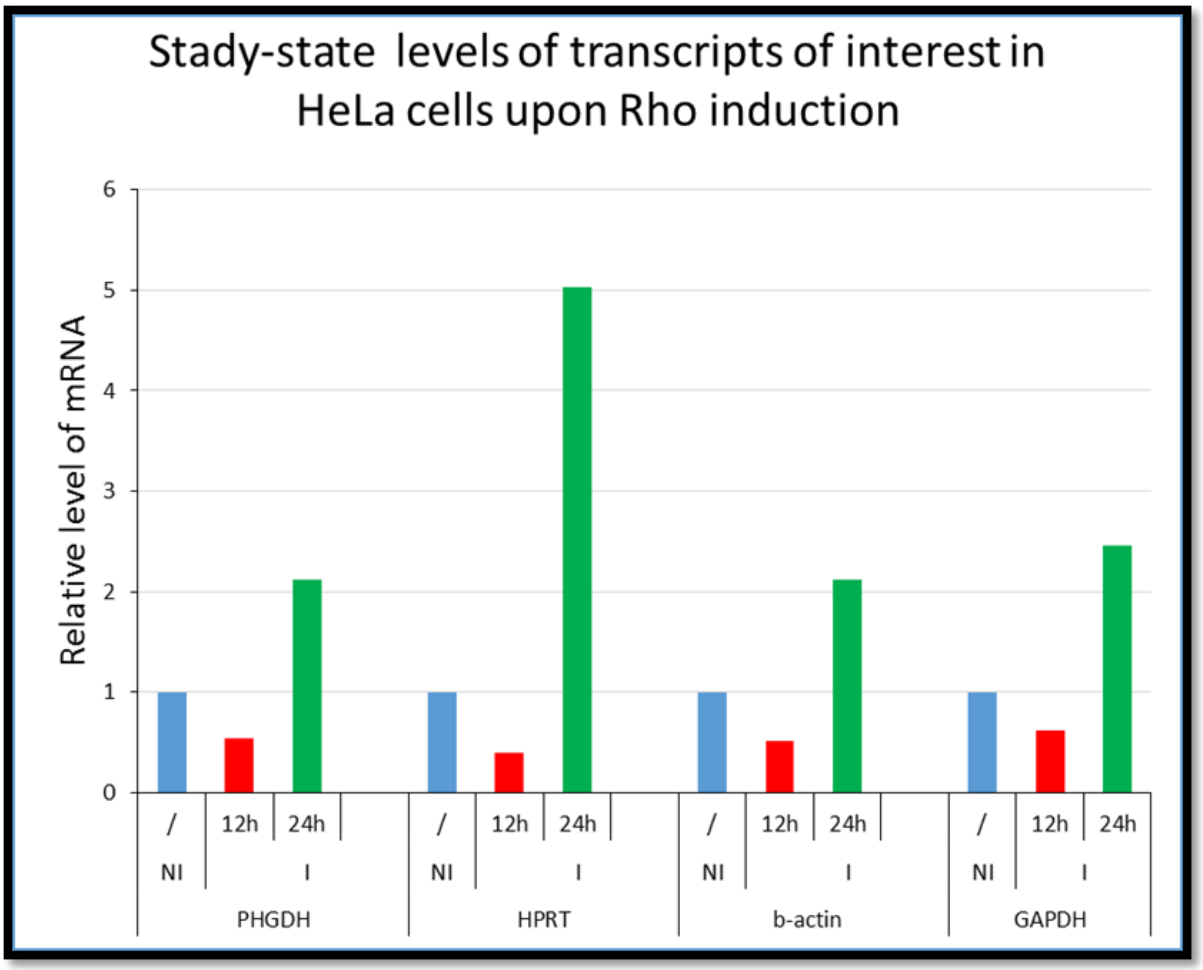


Figure 34: qPCR levels of 4 housekeeping genes after 12 (red) and 24 hours (green) of Rho induction in HeLa cells. Expression levels at 12 and 24h are normalized by gene expression in physiological condition (NI). Targeted genes are down regulated after 12h of Rho induction but are drastically up regulated after 48h of Rho induction.

The loss of expression of all 4 genes suggests that these transcripts are affected by Rho and are subsequently degraded by the degradation machinery. However, the restoration of expression after 24h of Rho induction is more cryptic. Overexpression after Rho induction was unlikely due to the typical transcript targeting of Rho. We hypothesized that this overexpression was a symptom of an accumulation of transcripts linked to the nuclear retention of aberrant mRNA via the transcript quality control and degradation system. Furthermore, we suggest that the accumulation of transcripts in the nucleus could be monitored via fluoroscopy.

2.3.3.2 Rho induction provokes the accumulation of mRNA and exosome exonuclease RRP6 in dot-shaped structures within the nucleus

Observation of Hela cells via FISH using dT-hybrided mRNAs showed that a 48h Rho induction led to the emergence of grain-like structures inside the nucleus. In addition, coupling FISH with RRP6 fluorescence showed that RRP6 was included within these structures. RRP6 is known to work with the exosome complex and to handle the degradation of ncRNA and aberrant transcripts. Considering both Cy3 coloration and RRP6 signal, we supposed that these structures encompassed or were made of ncRNAs and/or aberrant transcripts.

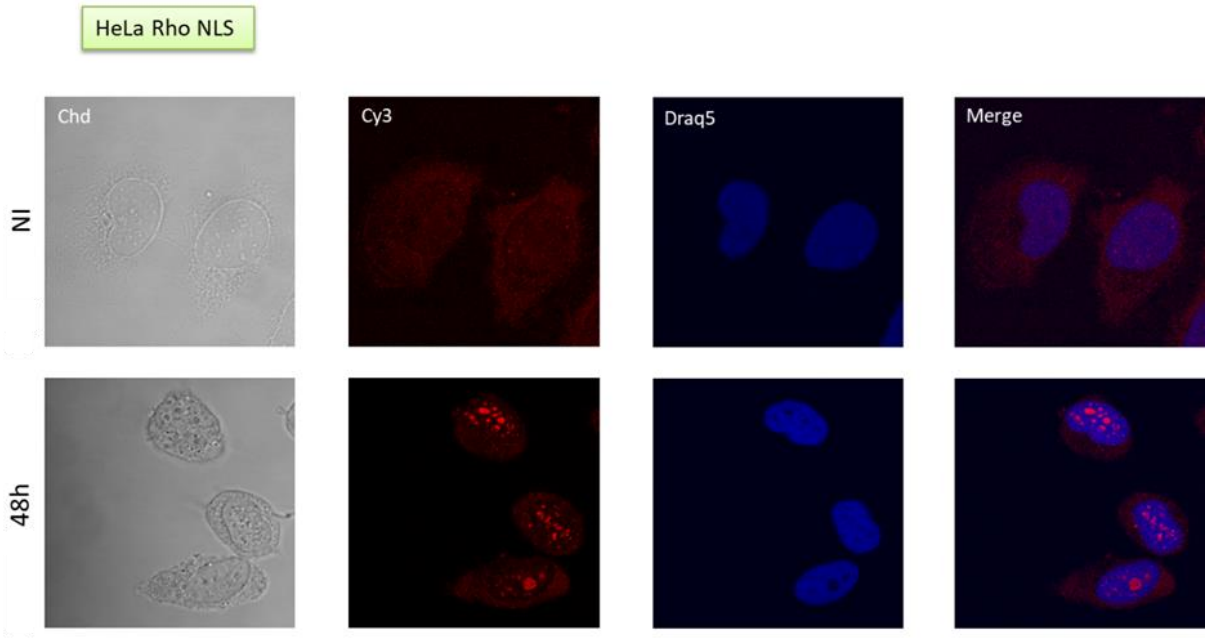


Figure 35: FISH microscopy targeting RNA (Cy3-red) with additional Draq5 nucleus coloration (blue) in HeLa cells with or without Rho induction (48h). The merge of both coloration reveals dots of RNA in the nucleus after 48h of Rho induction

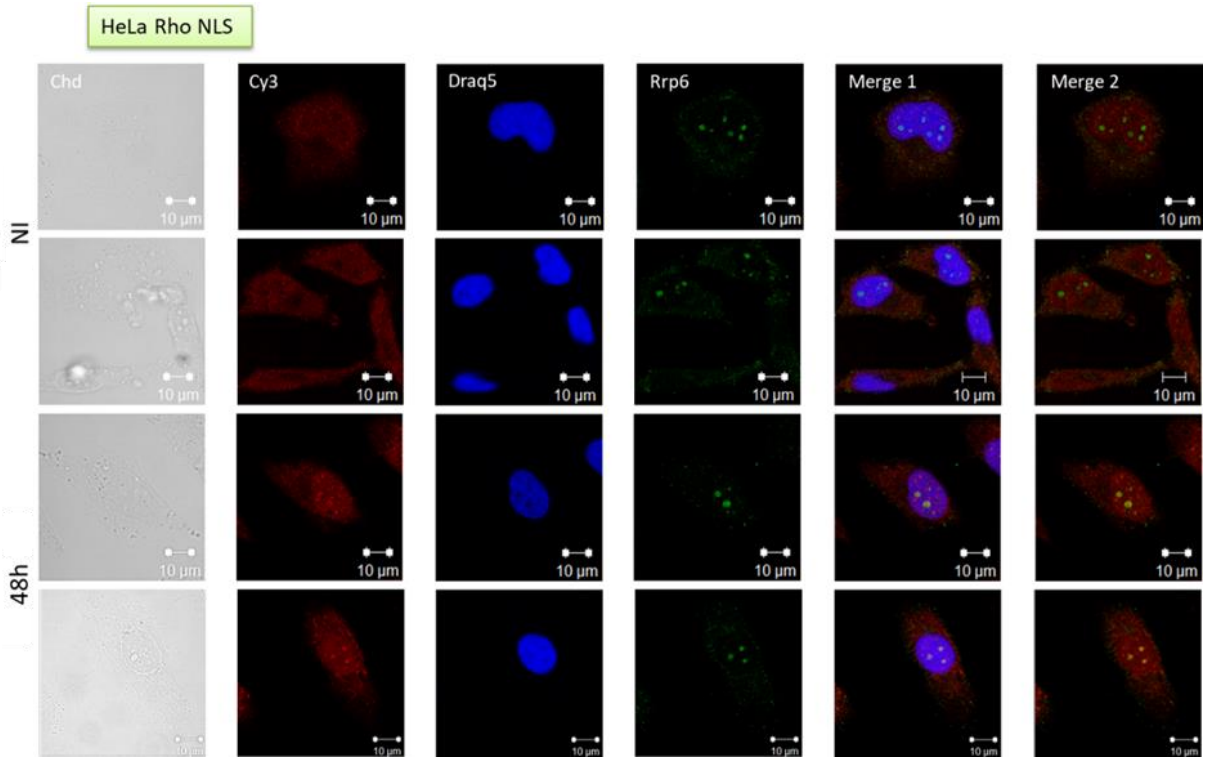


Figure 36: FISH microscopy targeting RNA (Cy3-red) with additional Draq5 nucleus coloration (blue) and Rrp6 (FITC-green) in HeLa cells with or without Rho induction (48h). Each condition was performed in duplicate. Merge of both colorations reveals dots of RNA in the nucleus after 48h of Rho induction that co-localize with Rrp6.

As suspected from our observation in qPCR (figure 36), we showed that transcripts are accumulated in the nucleus in dot-shaped structures after 48h of Rho induction (figure 37). Interestingly, this accumulation seems to be linked to RRP6 and probably the exosome as RRP6 co-localizes with the RNA enriched structure (Figure 38). RRP6 and the exosome are responsible for the degradation of a portion of Rho-induced aberrant RNA in yeast. If RRP6 also handles the degradation of Rho-induced transcripts in human, the recruitment of RRP6 in the dot-shaped structures could indicate that these structures are composed, at least partially, of Rho-induced aberrant transcripts. To ensure that RRP6 and the exosome were indeed responsible for Rho-induced aberrant transcript degradation, we conducted siRNA silencing of RRP6 and RRP40, a key component of the exosome.

2.3.3.3 Deletion of RRP6 and the exosome subunit RRP40 restore the expression of Rho-affected transcripts

Measurement of +Rho/-Rho differential expression of the selected genes showed a partial recuperation of the expression with the transfection of siRRP6. We also tested the effect of a deletion of a key component of the exosome, RRP40. Deletion of RRP40 is known to be more efficient in disrupting the exosome since it is part of the upper ring of the exosome core. This deletion led to a full recovery of the expression of the GAPDH gene and a more important expression of the HPRT and beta-actin genes than the basal level without Rho.

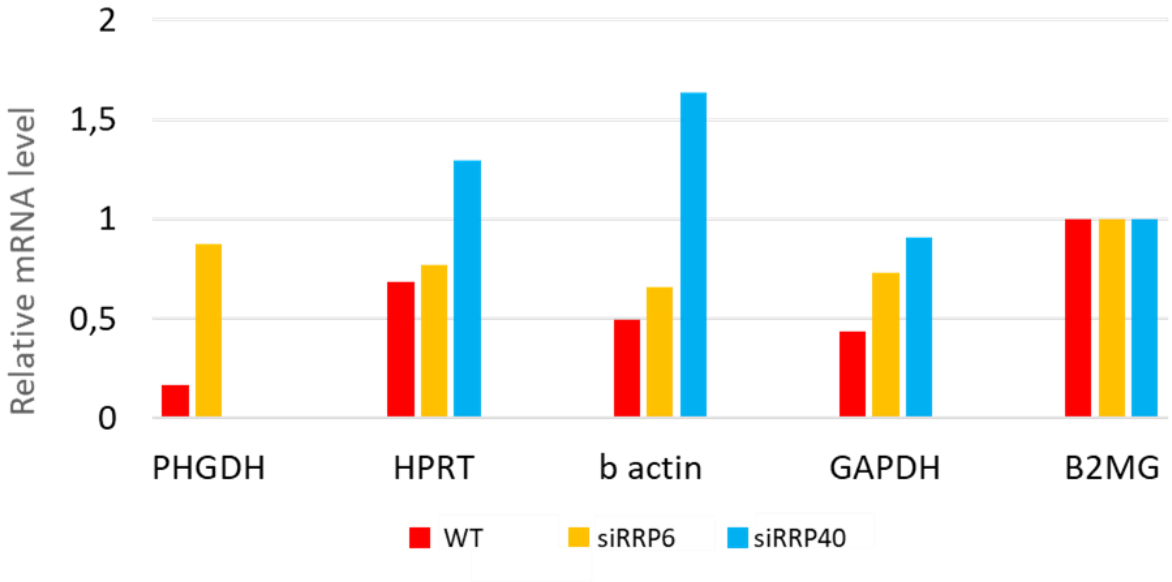


Figure 37: qPCR levels of 4 selected genes after 24h of Rho induction in HeLa cells in WT (red), transfected with siRRP6 (yellow) or siRRP40 (blue). B2MG serves as internal control and is not affected by Rho. Expression levels of the 4 genes are normalized by B2MG gene expression in each matching condition. Transfection with siRRP6 and siRRP40 leads to a recovery of the Rho effect. Deletion of RRP40 leads to a better recovery from the Rho effect than RRP6 deletion.

The restoration of expression after Rho induction in RRP6 silenced and RRP40 silenced strains indicates that RRP6 and the exosome are responsible for the degradation of Rho induced aberrant transcripts. Considering previous results from fluoroscopy, this indicates that the dot-shaped structures are composed at least partially of Rho induced aberrant mRNA. The nature of these dot-shaped structures and the role of RRP6/exosome in their formation remain unclear. However, an intra-

nuclear structure type called nuclear speckles presents similarities with what we observed with the Rho model.

2.3.3.4 Rho-induced aberrant mRNA accumulates within nuclear speckles

The grain-like structures (RNA dots) composed of exosome-linked proteins and RNAs are similar to the well-documented nuclear speckles. Since these speckles are mainly composed of aggregates of proteins like SC-35, which play a role in splicing, we checked the presence of SC-35 in the Rho-induced nuclear structures. Interestingly, Rho-induced grain-like structures co-localized with SR-35 nuclear speckles.

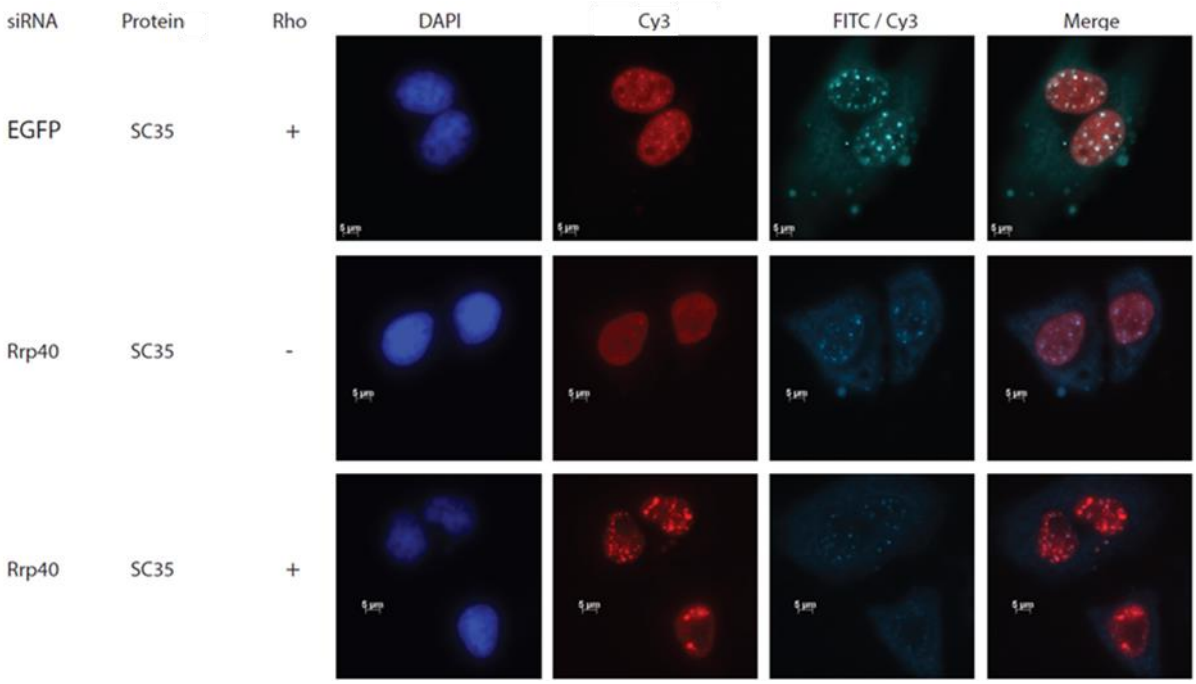


Figure 38: FISH microscopy targeting RNA (Cy3-red) with additional coloration of the nucleus with DAPI (blue) and SC35 (FITC-green) in HeLa cells with (+) or without (-) Rho induction (24h). In the two last rows, Rrp40 was silenced to deplete the exosome and block the degradation of the RNAs. The merging of both colorations revealed that dots of RNA in the nucleus after 24h of Rho induction co-localize with SC-35, a major component of the nuclear speckles.

SC-35 marking showed dot-shaped structures within the nucleus that we associate with nuclear speckles. This association seems to be Rho-independent as it can be observed in both the -Rho and +Rho conditions (Fig40, rows 1 and 2). We

observed that SC-35 co-localized with the RNA dots upon Rho induction (Fig40, row 3). Interestingly, the co-localization of SC-35 with the RNA dots seems to be exosome independent since the silencing of RRP40 and the subsequent depletion of the exosome does not suppress this co-localization nor the formation of speckles and RNA dots. The RNA dots, however, grew larger upon exosome depletion. As we thought that Rho induced a large concentration of RNA in the nucleus, we quantified this nuclear retention with qPCR, fractioned between nuclear and cytoplasm.

Quantification of the transcripts from 4 different genes in the cytoplasmic and nuclear fractions allowed us to monitor the nuclear retention of these transcripts due to Rho action. The obtained ratios show that a majority of the transcripts from these 4 genes are situated in the cytoplasm under physiological conditions. Upon Rho induction, the previous balance between cytoplasmic and nuclear transcripts is shifted. Transcripts from the HPRT, ACTB and GAPDH gene loci are mostly found in the nucleus after Rho induction, suggesting a retention of these transcripts within the nucleus, maybe in the identified speckles. Our results showed that a longer Rho induction increased the nuclear concentration of our transcripts of interest.

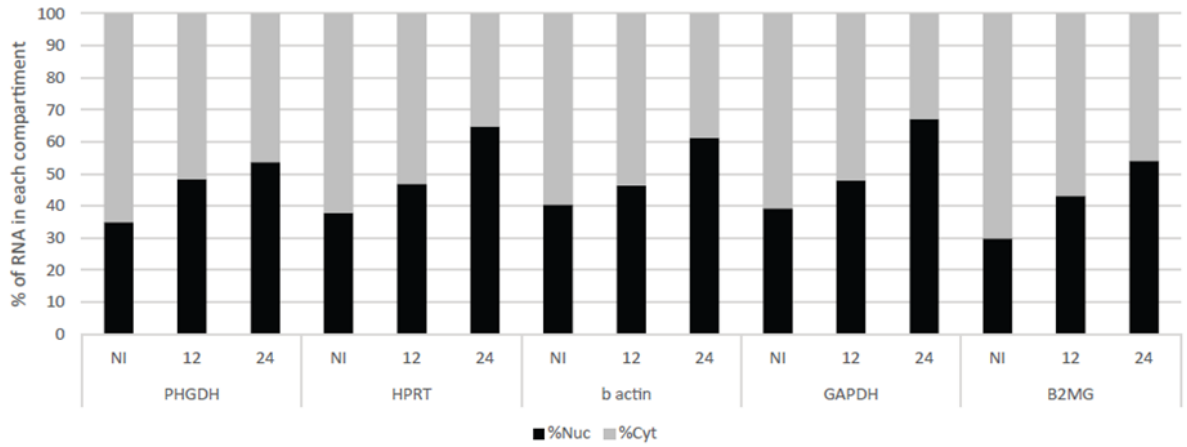


Figure 39: Relative distribution of the transcripts of 4 selected genes in the cytosol and nucleus, without Rho (NI), with a 12h or 24h Rho induction. In all cases, the induction of Rho led to a shift in balance between cytoplasmic and nuclear transcripts. The portion of nuclear transcripts for each gene increased with the presence of Rho, showing Rho-induced nuclear retention.

To gain a wider picture of the degradation process, we then aimed to measure precisely the effect of Rho on the entire transcriptome by performing an RNA-seq with the same conditions as previously stated. In addition, we separated the RNA-seq run into two parts, with cytosolic transcripts in one hand and nuclear transcripts in the other.

We aimed to study the expression levels of genes in both sub-cellular localizationq, to try to assess the retention of transcripts in the nucleus affected by the human quality control system (QC).

2.3.3.5 Identification of gene populations by bioinformatics

Using bioinformatics, we aimed to answer a few hypotheses arising from these previous results:

- Which transcripts are affected by Rho, and if so, which ones are retained in the nucleus?
- Why are transcripts retained in the nucleus? Link with NEXT and PAXT? Physical characteristics?
- Are SC-35 proteins sequestered within the Rho-induced nuclear speckles ?

More work needs to be performed to answer these questions. Thus, I will mostly present the bioinformatics results obtained at this time without considering their scientific impact as a whole but more to show the identification process of interesting genes.

Using a typical workflow for the preparation of the transcript counts, I then aimed to separate the different populations of transcripts depending on their kinetics upon Rho induction i.e. the variation in their numbers depending on the Rho induction time. The RNA-seq was designed as follows. Three experimental conditions are used: not induced (NI, also used as a t = 0), 12h and 24h of Rho induction. For each condition, the RNA extractions were fractioned to obtain either the cytosolic fraction or the nuclear fraction of RNAs, giving us a total of 6 different sample types. Each sample type was performed in biological duplicates to ensure the statistical meaning of the downstream bioinformatics analysis. Since most of the analyses are based on differential expressions, we calculated two sets of ratios: NI_12h for the differential expression between NI and 12h and 12h_24h for the differential expression between 12h and 24h. Each set of ratios was calculated in the cytosolic and nuclear fractions, giving a total of 4 ratio sets, which will be referred to as “conditions” in the rest of this result part.

We based our analyses on 19210 protein coding gene loci from Ensembl annotation, Read counts and transcript expression were calculated for this initial

population. Transcripts associated with 14573 genes were detected in the cytoplasmic fraction and 17670 in the nucleus fraction. From these two populations, transcripts from 13775 genes were found in both fractions and were kept for the following analyses. Finally, 4264 genes out of 13775 presented a variation in their expression in one or both conditions. This indicates that around 30% of the analyzed genes have their transcripts targeted by Rho or have their expression indirectly impacted by Rho induction. The variations in the expression of these genes were measured via differential analyses and gave a table with two variations: Non induced vs 12h and 12h vs 24h for each cellular compartment : cytosolic and nuclear. Overall, each gene is associated with 4 values, which are then used for downstream analyses.

The initial steps were to discover if each variable was related, as that would also mean that genes could present similarities between each other regarding their expression variation due to Rho induction. To test this, correlations between the expression variation of all the transcripts were calculated (figure 42 left). We observed no correlation between nucleus and cytosol gene ratios, whatever the Rho induction conditions. However, anti-correlation can be observed between NI/12h and 12h/24h ratios within the same subcellular localization. To be more precise, a large portion of transcripts is more or less expressed after 12h of Rho induction and this variation is then inversed after 24h of induction. This suggests a cohesive behavior of one or more groups of transcripts, but this behavior seems not to depend on the cellular compartment.

A clusterization by k-means was used to identify the sub-populations of transcripts. To this end, we focused not on the correlation between conditions but on the correlation between genes. Each gene was given a correlation factor with each other genes. Genes with similar variation of expression tend closer to 1, genes with opposite variation of expression tend to -1. Heatmap drawing of the ratio correlation suggested that several transcript groups coexisted (figure 42 right).

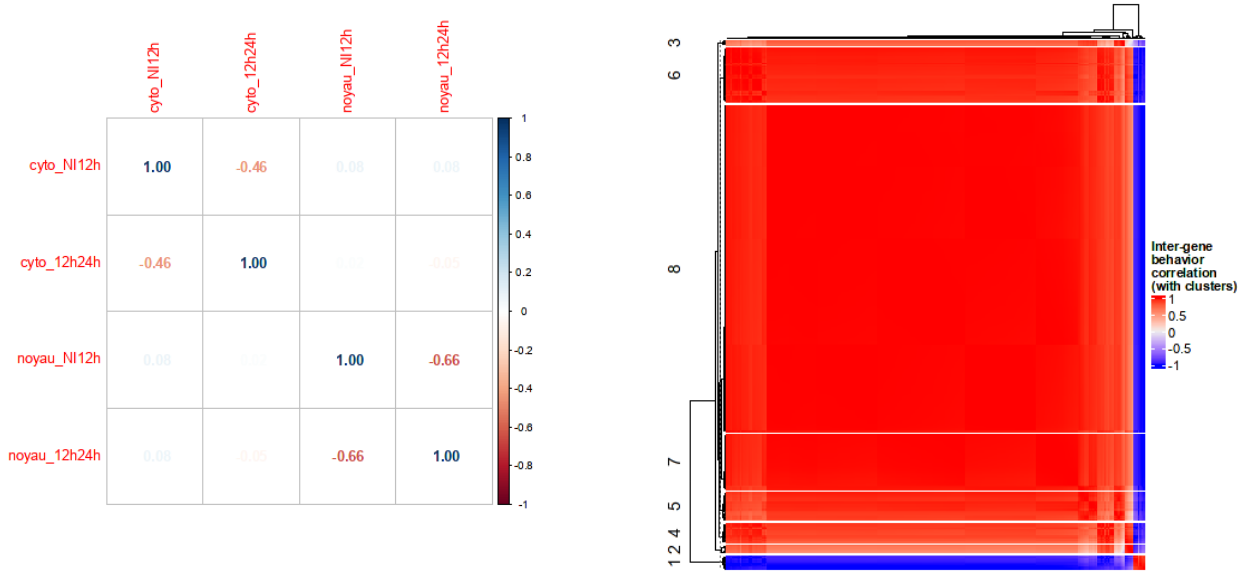


Figure 40: Correlation table between the logarithmic (log2) variations of expression of 4264 transcripts observed after 0-12h (NI12h) and 12-24 hours (12h24h) of Rho induction in the cytosolic (cyto) and nuclear (noyau) fractions. This table highlights an anti-correlation between the variations observed between 0 and 12h of Rho induction and the variations observed between 12 and 24h of Rho induction (left). Correlation heatmap of the expression variations measured in the cytosolic and nuclear fractions between 4264 genes. Each variation obtained for each gene was correlated with all the other genes to identify sub-populations. 8 sub-population were identified (right).

We considered 8 different subpopulations of genes. Using k-means clustering, 8 groups were formed from the total population of genes.

Cluster number	1	2	3	4	5	6	7	8
Population size (n)	1735	165	531	293	1117	110	2	311
Log2 cyto_NI12h	-0,33	-0,04	-0,08	0,06	0,28	-0,33	-9,45	0,14
Log2 cyto_12h24h	0,11	-0,08	0	-0,02	-0,13	-0,02	8,2	-0,03
Log2 nuc_NI12h	0	0,07	-1,24	0	0	-1,19	0	0,83
Log2 nuc_12h24h	0	-0,97	1,41	1,18	0	0,02	-1,03	-0,01

Figure 41: Table representing the expression variation and population size of the 8 sub-population of transcripts identified by clustering. First two rows are the cluster numbers and their population size (n). 4 last rows are the averaged expression variation (log2) of the transcripts from each cluster in the cytoplasm (cyto) and nucleus (nuc) between 0 and 12h (NI12h) and between 12 and 24 hours of Rho induction (12h24h).

GoTerm analysis did not reveal any common pathway within clusters, indicating that Rho affected transcripts independently of their biological finality. Additional steps can be considered. Since we wished to identify the transcripts retained in the nucleus, expression levels within both the nucleus and the cytoplasm were compared but the use of transcript counts instead of variation ratio would be more pertinent to quantify the nuclear retention phenomenon.

The next step would be to explain why a given transcript is retained in the nucleus. A first and rather quick option would be to check some physical properties of the transcripts such as their length, GC content, and number of exons/introns. By building a table with several variables, an ACP could be performed to determine if one or more physical characteristics could separate genes from one cluster from another. A second and more laborious option would be the use of an external dataset of RNA-seq or equivalent techniques to know which transcripts interact with NEXT or PAXT complexes. A good source would be the HTS resources cited in [194, 241].

Finally, we know that a major protein composing the speckles is SC-35, a protein implicated in the splicing of RNAs. We can hypothesize that the retention of these proteins in the speckles could impede splicing. We could measure the intron-skipping phenomenon to check this hypothesis. Good experimental programs are RMATS for the core processing and the R package maser for the post-processing. I will discuss below further options for the continuation of this project.

3. Discussion

The discussion part will be subdivided into 4 parts. The first one deals with the precedent ongoing results. The second one will add a few words about the two publications written during my PhD. I will finish by discussing how the discoveries made during my PhD integrate with the whole mRNP biogenesis process.

3.1 Rho induction in HeLa cells provoke the formation of RNA-filled nuclear speckles

Although our last project is not complete, our observations allow us to draw a model of the action of Rho in human cells. First, the predicted action of Rho on mRNAs

was similar to what is documented in yeast, as the action of Rho led to the disruption of the correct production of a portion of the population of mRNPs. A measure of housekeeping gene expression showed a decrease with Rho induction. Additionally, a relatively large number of genes have their transcripts affected by Rho, as identified via bioinformatics. Go-Term analysis showed that the implicated transcripts belong to various biological pathways. This suggests a non-specific interaction of Rho with the transcripts, which can be expected considering the nature of Rho-RNA interaction, which only depends on RNA structure and GC-content. However, the induction of Rho revealed an unexpected mechanism linked to the induction of aberrant mRNA in the nucleus. Longer induction of Rho led to the apparent restoration of the expression of Rho-affected genes. Moreover, supplementary observations of cells upon long Rho induction showed that the transcripts were accumulating within the nucleus in dot-shaped structures visible upon probe-tagging of polyadenylated RNA with fluorescence microscopy. The accumulation of transcripts in the nucleus was validated using fractioned qPCR, allowing us to calculate the repartition of the transcripts in the cell between the cytoplasm and nucleus. The nuclear retention of aberrant transcripts upon mRNP biogenesis disruption is largely documented in yeast and humans [242-244]. Retained transcripts are then degraded via the exosome and its associated co-factors. Curiously, in humans, it seems that Rho-induced aberrant transcripts are not only retained in the nucleus but also accumulate abnormally. In yeast, the exosome exonuclease Rrp6 usually handles the degradation of Rho-induced aberrant transcripts. We sought to verify if RRP6 (human Rrp6) was implicated in the degradation of Rho-induced aberrant mRNP in human. By using immunofluorescence, we observed a co-localization between mRNA dots and RRP6, suggesting that it was indeed implicated at one point in the accumulation observed, although it is unclear if RRP6 and the exosome are the main reason that tagged mRNP are packed in dot-shaped structures.

The nuclear dot-shaped structures were already documented to appear in pathologies involving aberrant mRNPs, such as Huntington disease. In this pathology, HTT transcripts are aberrant due to an abnormal number of CAG repeats within the HTT coding gene. These CAG repeats lead to an agglomeration of the afflicted mRNAs and their inability to be exported outside of the nucleus [245]. These aberrant mRNPs assemble in dot-shaped structures, which are similar to the ones we observed upon

Rho induction. These dot-shaped structures within the nucleus are called nuclear-speckles. One of their primary components is the protein SC-35, which is usually implicated in splicing. By combining RRP6 and SC-35 immunofluorescence with poly(dT) probing, we discovered that mRNA, RRP6 and SC-35 were all part of the dot-shaped structures observed after Rho induction. This result confirms that the structures we observed are nuclear speckles with a high concentration of mRNA due to Rho induction. If Rho works in a similar way in human than in yeast and remove packaging proteins like the THO/TREX complex, we can imagine two non-exclusive mechanisms explaining our observations.

1) A large variety of mRNP transits via the nuclear speckles (NS) mostly, but not only for splicing [246]. For example, nuclear speckles are known to be an important component in the nuclear export of intron-less transcripts [247]. This is achieved with the recruitment of the human TREX complex onto the intron less mRNA while localized within the nuclear speckles. The induction of Rho could influence TREX recruitment on intron-less transcripts and lead to the retention of the affected transcripts in the NS. In addition, the decay of intron-less mRNA by the exosome is linked to the nuclear speckles via SRSF3, a component of the speckles [248, 249]. If intron-less mRNAs are made un-exportable due to Rho removing TREX, we can reasonably suppose that the exosome could be recruited on site, hence explaining the co-localization of SC-35, mRNA and RRP6 and the accumulation of mRNA within the NS. Bioinformatics analysis of the number of exons of transcripts affected by Rho and retained in the nucleus could help us figure out if what we observed are only intron-less transcripts or not within the nuclear speckles. However, since Rho has low specificity when it comes to mRNA targeting, it is likely that intron-containing transcripts are also affected.

2) In bacteria, Rho terminates transcription by ejecting the polymerase from the transcripts. It is also documented that the translocase activity of Rho is strong enough to remove streptavidin from the transcripts and we hypothesize that it is strong enough to remove some if not most packaging proteins from various mRNP in yeast like TREX. Although we do not know if Rho translocase activity can generate enough mechanical stress to remove

the spliceosome, we can reasonably suggest that the action of Rho could at least hinder the correct splicing of its targets. For example, Rho action could disrupt the early assembly of the spliceosome or block exon ligation after intron removal. The affected transcripts could then be considered aberrant. There are several possibilities to explain how they are retained within nuclear speckles, which will be explored below.

Two questions arise from these hypotheses. The first point to be discussed is how and why mRNA, aberrant or not, would be retained in the nuclear speckles. Nuclear speckles are tightly linked to the splicing processes, and are well known to be made of a large amount of splicing proteins and snRNAs. However, it is still debated if the splicing occurs co- or post-transcriptionally, or even both. Similarly, the functions of the nuclear speckles are not fully understood, and while it is admitted that they serve as splicing factor storage centers, it is not clear how these factors are recruited to splicing sites. However, recent works suggest that nuclear speckles could be physically implicated in splicing by being key structures in which exons would be able to bind to each other easily during the late phase of splicing [246]. A simpler model suggests that the density and subsequent enrichment for splicing proteins of nuclear speckles made them perfect to perform splicing operations reliably [250]. In both cases, instead of having the splicing factors recruited from the speckles to the splicing sites, the RNA is routed to the nuclear speckles. To fit with the co-transcriptional splicing hypothesis, it is possible that exon-intron cleavage occurs transcriptionally and exon-exon ligation occurs near or within the speckles.

The second point to discuss would be the reason that mRNPs accumulate in nuclear speckles. The accumulation of mRNP observed in the nucleus could be the consequence of a degradation machinery overload or a “traffic jam” near the nuclear speckles. It would mean that NS either allow storage of aberrant mRNP as observed in Huntington disease or are a crucial actor for mRNP processing and/or export, which is already the case for some mRNAs and is suggested in some previously cited studies. If the degradation machinery is overloaded, the drop of mRNP within the nucleus after 12h of induction shows that the degradation machinery is indeed mobilized to cope with Rho induced aberrant transcripts and the 24h accumulation could be a sign of exhaustion of this machinery. If there were a “traffic jam” near the NS, aberrant mRNA

would be retained in the nuclear speckles waiting for degradation, effectively blocking correctly formed transcripts from being spliced. As the delay for splicing grows dire, unspliced transcripts and aberrant transcripts would accumulate in the NS until the jam becomes dire enough to be visible with fluoroscopy. Subsequently, the accumulation observed in the nucleus could also be partially due to splicing defects. Transcripts retained for degradation within the nuclear speckles could sequester splicing proteins along with them. The lack of available splicing proteins could provoke splicing defects of some mRNAs, which would then be considered aberrant, resulting in more and more transcripts being targeted by the exosome until saturation. This saturation of the exosome machinery and the perturbation of the splicing mechanism could lead to the observed accumulation and would explain the co-localization of RRP6, SC-35 and mRNA, while explaining the lack of accumulation after only 12h of Rho induction.

Some of the previously cited hypotheses may also be explored further by using bioinformatics. For example, validation of our splicing defect hypothesis due to SC-35 retention with aberrant mRNA would require the quantification of intron-skipping events between WT and Rho-induced conditions. We could also validate if intron-less transcripts are preferentially retained in the nucleus. This would be accomplished by drawing a map of general characteristics (gene length, exon/intron number, GC content,…) of Rho-affected transcripts and determining if some of these characteristics could be overrepresented in some clusters. This type of application could be broadened to include exosome cofactor preferences (NEXT or PAXT) using previously published datasets [194, 241].

Additionally, our analysis could be expanded to include ncRNAs. The quantification of ncRNAs, especially ncRNAs like PROMPTs could provide insights on the regulation of ncRNA transcription. As identified in yeasts with exosome inactivation [25], the induction of aberrant mRNPs could divert a portion of the exosome targeting systems from ncRNA loci to mRNA coding ones, leading to an overexpression of ncRNA. Concretely, we would observe a larger number of specific families of ncRNA (like PROMPT) combined with more frequent termination defects at non-coding gene loci upon Rho induction.

3.2 Tho2 moonlights in yeast co-transcriptional mRNP quality control by targeting aberrant mRNPs to Rrp6

Thanks to the democratization of HTS techniques and more globally the “omics” field of research, we were able to evaluate the functioning of the QC degradation system as a whole. By monitoring the recruitment of key components of the aberrant mRNP targeting and degradation system, we highlighted the importance of Tho2, a protein canonically involved in mRNP packaging. In the canonical system, Tho2 associates with 3 other proteins to form the THO complex. This complex is responsible for the early packaging of the mRNP and is required for its export. The interaction between Tho2 and mRNPs has different outcomes depending on how Tho2 is involved in the THO complex. When Tho2 interacts with the transcript alone, it triggers the recruitment of the degradation system i.e. the exosome. Whether the recruitment is direct or depends on cofactors is unclear, as other factors are also capable of recruiting the exosome, such as Isw1 [251]. Notably, Gbp2 a protein already known to be implicated in splicing and quality control is able to link Tho2 and Rrp6 [252]. We can suppose that Tho2 and Gbp2 can interact when both are monomeric. The interaction site of Tho2 with Gbp2 could be masked by other THO subunits when Tho2 is in the complex. Alone, this site could be accessible to Gbp2 and the Tho2-Gbp2 interaction could lead directly or indirectly to the recruitment of Rrp6/the exosome. The involvement of complexes in mRNA degradation that are typically involved in other pathways of transcription and/or degradation is not new. While the NNS complex is a key component in the processing and degradation of some ncRNA, it also plays a role in the degradation of aberrant mRNP [25, 253].

It is clear that our contribution does not conclude the story of the relationship between Tho2 and Rrp6. Apart from a possible physical link between Rrp6 and Tho2, which remains to be demonstrated, many proteins could be a bridge between Rrp6 and Tho2. Indeed, any cofactor of Rrp6 or Tho2 could be considered a potential candidate. Investigating these potential candidates and Rrp6 – Tho2 interactions is a comprehensive and ambitious project that could use powerful techniques such as IP-MS (immunoprecipitation-mass spectrometry) for screening followed by Co-IPs of selected candidates. Techniques combining ChIP and MS could also be used to determine the functional interactions of Tho2, Rrp6 and their eventual cofactors with

the chromatin. In this case, the loci enriched with Tho2 upon Rho induction identified in this study could be reused to validate the experiment. A more interesting and time-saving way to perform such an amount of analysis would be to use Co-fractionation/mass spectrometry (CF/MS, [254]) with and without Rho induction to list any Rrp6-Tho2 complexes, and enumerate their co-factors.

Another interesting aspect to explore could be the mechanisms revolving around the dimerization of the THO complex. The mechanistic model proposed for the THO/TREX dimer suggests that the complex could travel along the nascent RNA thanks to the ATPase activity of Sub2 [91]. Although we lack evidence of a mobile complex, it could have a major impact on how we conceive THO activity and recruitment. For example, considering that THO/TREX is recruited to prevent R-loop formation, its ability to travel along the transcripts could mean that it could follow the polymerase instead of being fixed structures on the RNA. Another example given by the authors is that mobile TREX complex could disrupt deleterious secondary structures. We assumed that Rho disrupts the dimers, as it does to the complex. An alternative hypothesis, though more convoluted, would be that Rho hinders the translocation of the complex, potentially leading to a partial disassembly of the complex mimicking a THO complex disruption. Nevertheless, the possibilities that THO dimerization bring to the field deserve more research, with or without Rho. The dimerization of THO/TREX could be further explored using the structural elements highlighted in *Schuller et al.* and site-directed mutagenesis. However monitoring the monomeric or dimeric state of THO on the RNA might be a challenge. One ambitious option could be to modify Tho2 to harbor a moiety of a tag flanked by a poly-His tail. Upon dimerization, the tag moieties would be close enough to assemble into the full tag by using a method derived from a biochemical study, in which the His-tag are used to from dimers [255]. Two problems among many is that this approach requires that the histidine bond could withstand the interaction with the antibody and that the dimerization is not mediated via the His tails.

Finally, this study could also be transposed to humans. The THO complex was also identified in human, in which it is composed of 6 subunits. While 4 of them are counterparts of the yeast proteins Hpr1, Tex1, Tho2 and Mft1, two of them do not match any yeast THO complex subunits [95]. It would be interesting to study if the human

Tho2 harbors the same specificities that we found in yeast. This could be conducted using a method with Rho and ChIP-seq, or a more traditional method using RIP-seq and gene-silencing of THO subunits.

3.3 Yeast RNA exosome activity is necessary for maintaining cell wall stability through proper protein glycosylation

Combining the bioinformatics approach with a more traditional experimental aspect in biology allowed us to explore more deeply some mechanisms of the exosome and reveal the consequences of exosome depletion in transcription regulation. It is known that the deletion of Rrp6 leads to the downregulation of a large number of genes [25]. Deletion of Rrp6 in yeast leads to a growth defect and a temperature sensitivity. This temperature sensitivity is reduced by the addition of an osmotic stabilizer to the culture medium. Analyses using bioinformatics allowed us to identify a few genes implicated in the integrity of the cell wall. The expression of these genes was then quantified by qPCR to validate the results. By analyzing the expression of antisense and promoter proximal ncRNA, we found that one particular gene (PSA1), which is strongly downregulated upon Rrp6 deletion, had its promoter overlapped by a highly expressed CUT in the same conditions. Our work shows the importance of pervasive and antisense transcription for the regulation of gene expression, following in the footsteps of other recent works [22, 256].

3.4 Conclusion

During my PhD, I explored the mechanism underlying the targeting and degradation of aberrant mRNPs by using Rho as an mRNP biogenesis disrupter. My work benefited from the use of Rho as a model to study mRNP quality control in yeast that was initiated by Dr. Rachid Rahmouni's team just over a decade ago [217]. Recent work broadens our research spectrum by using bioinformatics to assess the effect of the Rho model on the entire population of transcripts [25]. In continuity with the approach developed by Dr. Kévin Moreau, I adopted bioinformatics to study the THO complex and later the effect of Rho in HeLa cells. Like most scientific work, my work on the THO complex raised as many questions as it answered. We have shown that Tho2 and Rrp6 are indeed linked, at least functionally. However, it is still unclear how

and by which proteins this connection is made, if not by Rrp6 and Tho2 themselves. The entire system that controls aberrant mRNA degradation also remains to be further defined. While the degradation actor, i.e. the exosome, is well known, the understanding of the targeting of aberrant mRNA is still obscure. In the literature, several proteins that are key players in transcript processing are involved in the retention or targeting of aberrant transcripts [257, 258]. This could imply that there is no dedicated targeting system but that it is fully integrated into the entire mRNP biogenesis pathway. In this model, some players involved in mRNP biogenesis, such as Tho2, could serve as a flag for deletion in the event of mRNP assembly failure. Each step of biogenesis, from packaging to export and including splicing, capping and polyA tailing could have its own flagging protein(s) and perhaps its own degradation pathway. For example in yeast, Rat1, an exonuclease involved in transcription termination and rRNA/snoRNA processing, is also responsible for the degradation of aberrant HXK1 transcripts upon Rho induction and has been suggested as a minor degradation pathway of mRNA in the nucleus [216, 259]. In addition to this hypothetical model, recruitment of ncRNA processing factors such as Nrd1 to mRNA coding loci upon mRNP biogenesis provides a second layer of protection against aberrant transcripts that expose similarities to ncRNA i.e. with accessible Nrd1 interaction sites. Considering the results obtained with the Δrrp6 strain and other work on ncRNA transcriptional defects, the displacement of ncRNA processing factors to mRNA coding loci could have an effect similar to the deletion of Rrp6 [253]. Lack of ncRNA processing factors is known to lead to termination errors, and it has been shown that such impairment could hinder the transcription of mRNA coding genes downstream of the ncRNA gene in a phenomenon known as transcriptional interference [22]. The overall effect of the ncRNA termination defect and the disruption of transcription of the coding genes by the deletion of Rrp6 results in growth defect in yeast. Hypothetically, this growth defect could be useful for cell survival in cases of aberrant mRNA biogenesis, especially if the origin of the aberration is due to genomic damage. If flag proteins like Tho2 could be a first line of defense against mRNP assembly errors by quickly identifying aberrant mRNA, the recruitment of ncRNA processing factors could be a second line of defense in case of a critical situation when a large number of aberrant transcripts are produced. The effect that displacement of ncRNA processing factors has on the transcription of coding genes and physiological growth could be an

evolutionary way to prevent too much aberrant mRNA from being produced and to give more time to the genomic reparation system to operate on eventual damage to DNA.

A large series of experiments could be considered to answer this ambitious hypothesis. Using Rho as a model for aberrant mRNP biogenesis seems ideal to study both Rrp6, THO and possible cofactors without removing any protein that might play a role in aberrant mRNP targeting and degradation. Identification of proteins linked to the chromatin upon Rho induction could be accomplished using a recently developed approach that employs mass-spectrometry for chromatin-bound proteins, and also uses bioinformatics to identify complexes from a large amount of data [260].

4. Résumé en français

Ce résumé reprend le manuscrit en anglais ci-dessus. Afin de répondre au cahier des charges, le contenu de ce résumé est fortement tronqué par rapport au manuscrit en anglais. Bien que compréhensible, il manque donc de contexte et d'informations, tout particulièrement pour la partie résultat, qui reprend les publications produites pendant ma thèse. J'invite donc fortement le lecteur à lire le manuscrit en anglais.

1. Introduction

La transcription est le processus biologique menant à la production d'ARN à partir d'une matrice d'ADN. Ce processus est présent chez tous les organismes vivants, qu'ils soient procaryotes ou eucaryotes, avec quelques variations fonctionnelles entre ces domaines. La transcription étant le mécanisme clé entre notre code biologique, l'ADN, et les acteurs physiques de notre métabolisme, les protéines, ce processus a été étudié en profondeur au cours des 60 dernières années. La première preuve de l'activité de l'ARN polymérase a été découverte en 1959 par Weiss et Gladstone dans le foie de rat [1]. Depuis, les recherches sur le processus de transcription se sont poursuivies. Trois formes de polymérase ont été découvertes en 1969 chez les eucaryotes [2]. Ces trois complexes ont été nommés PolI, II et III. De grands panels de travaux biochimiques ont été réalisés pour découvrir le rôle respectif de ces 3 formes. Si tous ces complexes assurent la transcription de l'ADN en ARN, leur substrat ADN et l'ARN produit diffèrent. Plus particulièrement, PolII transcrit les gènes codant pour les protéines en ARN messager (ARNm), mais également un large panel de transcrits communément appelés ARN non codants (ARNnc), comme les transcrits cryptiques instables (CUT) et les transcrits stables non caractérisés (SUT) chez la levure. Chez les procaryotes, une seule forme de polymérase est chargée de produire toutes les sortes d'ARN [3-5].

Un ARN (acide ribonucléique) est un polymère de quatre types différents de nucléotides. Les nucléotides sont disposés dans une séquence qui détermine les caractéristiques de l'ARN, depuis sa protéine codée (pour les ARNm), sa capacité de survie (queue polyA, contenu GC), jusqu'à ses capacités catalytiques (activité catalytique de certains ARNnc). Les ARN sont générés à partir d'une matrice d'ADN.

La matrice d'ADN, le génome, est distribuée sur des composants appelés chromosomes (16 chez la levure, 46 chez l'humain, jusqu'à 4 chromosomes chez les procaryotes [6]). Les chromosomes hébergent des locus spécifiques appelés gènes, chacun codant pour un ARN. Certains gènes peuvent coder pour l'ARN messager (gènes codants) ou pour l'ARN non codant (gènes non codants).

1.1 ARNm

Le rôle principal de l'ARN messager est de transporter l'information génétique contenue dans le génome jusqu'au ribosome dans le cytosol pour produire les protéines.

L'ARNm possède trois domaines distincts dont la longueur varie. La partie principale est le cadre de lecture ouvert (ORF), qui contient la séquence d'ARN, qui sera traduite en protéine. L'ORF est flanqué de deux régions non traduites (UTR). Le 5' UTR commence à l'extrémité 5' du transcrit jusqu'au premier nucléotide de l'ORF, tandis que le 3' UTR commence au dernier codon STOP de l'ORF et se poursuit jusqu'à l'extrémité 3' du transcrit. Comme leurs noms l'indiquent, ces régions ne seront pas traduites et n'influenceront pas la séquence de la protéine résultante.

Alors que la sous-famille d'ARNm est cruciale pour le codage des protéines, d'autres sous-familles d'ARN jouent un rôle essentiel. L'ARN non codant est une sous-famille d'ARN qui, contrairement à l'ARNm, ne produira aucune protéine, bien que des études récentes indiquent que des peptides peuvent provenir d'une forme de traduction des ARNnc inexplorée [20]. Ils sont répartis en différentes classes en fonction de leur rôle, de leurs taux de dégradation et, pour certains, de leur localisation dans le noyau. Certains ARNnc ont été identifiés dans plus d'une sous-famille différente car des équipes distinctes les ont découverts avec des techniques différentes. Dans le cadre de ce résumé, je n'explorerai que deux sous classes de longs ARN non codants identifiées dans la levure *Saccaromyces Cerevisiae*, à savoir les CUTs et les SUTs.

1.2 Exemples d'ARN longs non codants chez la levure : Transcripts instables cryptiques (CUT) et Transcripts stables non caractérisés (SUT)

CUT peut se traduire par « transcrits cryptiques instables ». Ils sont produits par la polymérase II (PolII). Cette famille d'ARNnc reste dans le noyau et est rapidement dégradée après ou même pendant leur transcription [21]. Ils sont générés à partir d'une séquence intronique , de loci antisens et également par « transcription pervasive» [22]. Leurs rôles ne sont pas très clairs mais certains travaux indiquent qu'ils peuvent moduler négativement le taux de transcription des gènes proches en provoquant une interférence sur la transcription, une modification des histones ou une méthylation de l'ADN [23]. Le surenroulement lors de la transcription des ARNnc et la présence des acteurs de la transcription pourraient altérer l'ouverture de la chromatine et/ou le recrutement de PolII dans les gènes voisins [24]. La principale hypothèse est que la transcription d'un ARNnc bloque ou empêche également toute transcription du brin opposé, qui peut contenir des locus codants, réprimant le gène. Une autre hypothèse propose que les ARNnc joue un rôle sur la régulation de facteurs de transcription, agissant comme des « zone de stockage » pour ces régulateurs tant qu'ils ne sont pas nécessaires sur des loci plus importants. Ces ARN subissent des modifications post-transcriptionnelles avant leur dégradation comme le coiffage et la queue polyA , bien que le dernier processus soit différent entre l'ARNm et le CUT [25].

SUT signifie transcriptions stables non caractérisées. Leur rôle semble assez similaire à celui du CUT, mais ils sont plus stables car leur voie de dégradation est différente. Ils sont encore mal compris. Certaines études indiquent que leur dégradation pourrait être géré de la même manière que celui de l'ARNm. [27].

Après avoir présenté les différents types d'ARN sur lesquels j'ai travaillé, je vais parler de la maturation des ARN et plus particulièrement des ARN messagers par des protéines d'empaquetage.

1.3 Formation de mRNP et maturation de l'ARNm : exemple du complexe THO

Pour être fonctionnel, un ARNm naissant subit de multiples modifications. Ces modifications peuvent être apportées directement sur l'ARN, comme une coiffe en 5' ou l'épissage d'une portion du transcrit, ou par ajout de protéines qui auront des rôles

spécifiques dans la protection des transcrits, le contrôle qualité ou l'export nucléaire par exemple. Le résultat d'une association entre un ARNm et ces protéines est appelée ribonucléoprotéine messagère (mRNP). Un exemple de ces cofacteurs est le complexe THO qui a été mon principal sujet d'étude. Dans cette partie, je vais disserter sur ce complexe, sa structure et ses rôles.

1.3.1 Structure

Décrit il y a deux décennies chez la levure, le complexe THO (THO) est un facteur clé dans la biogenèse des mRNPs [90]. Il est composé de quatre protéines principales, Tho2 (180 kDa), Hpr1 (90 kDa), Mft1 (45 kDa) et Thp2 (30 kDa). Tex1 (47 kDa), même si elle n'est pas critique pour la formation et l'intégrité des complexes, est souvent considéré comme une cinquième sous-unité non essentielle. L'analyse Cryo-EM a montré que le complexe THO se dimérise lors d'une reconstruction in vitro. La dimérisation semble être médiée par les sous-unités Thp2 et Mft1 [91]. La régulation de la formation du complexe THO pourrait être réalisée via Hpr1, car elle peut être ubiquitinée par la voie de l'ubiquitine ligase Rsp5 [93].

Ce complexe s'associe ensuite aux protéines de type SR Sub2 (UAP56 chez l'homme) et Yra1 (Ref/Aly chez l'homme) pour former le complexe TREX. Le complexe THO-Sub2 est toujours un dimère mais il est supposé qu'une seule copie de Yra1 interagit avec le dimère THO-Sub2. L'interaction de THO avec Tex1 et Sub2 est médiée par Tho2 et Hpr1. [95].

L'interaction de THO avec son substrat ARN est gérée par Tho2. L'extrémité C-terminale de Tho2 (Tho2-CTD) est faiblement structurée et principalement composée de résidus chargés positivement. Les alignements de séquences multiples n'ont révélé aucun motif conservé ou identifié au sein du Tho2-CTD, mais sa troncature a entraîné une perte de l'interaction Tho2-ARN. Bien que cela ne soit pas critique pour la stabilité du complexe THO, la perte du Tho2-CTD entraîne une altération de l'expression génique similaire aux phénotypes observés lors d'une délétion de Tho2 et entrave l'interaction de THO avec la chromatine [96].

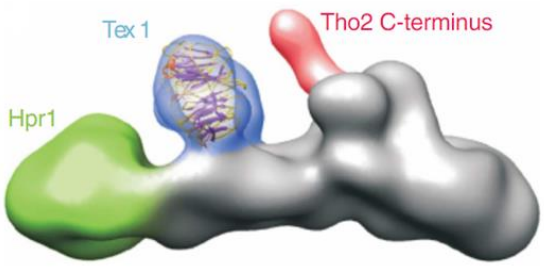


Figure R1 : L'extrémité C du Tho2 est placée sur une saillie au sommet du THO. Tiré de [96].

Deux protéines de type SR sont également connues pour interagir avec THO tout en jouant un rôle dans l'épissage : Gbp2 et Hrb1. Elles sont connues pour transporter le mRNP épissé vers la membrane nucléaire et Gbp2 interagit avec le domaine C-terminal de Tho2 pour son chargement sur le transcrit [94, 97]. Dans le cas où l'épissage échoue, Gbp2 et Hrb1 recrutent le complexe TRAMP, un composant du processus de contrôle de qualité (QC) des mRNPs, conduisant à une dégradation du transcrit [98].

1.3.2 Recrutement et rôle

Le complexe THO (THO) est recruté de manière co-transcriptionelle sur la chromatine en interagissant avec le domaine C-terminal (CTD) de l'ARN PolII puis avec les transcrits naissants [99]. Le recrutement de THO sur l'ARN PolII dépend de l'état de phosphorylation du CTD. Chez la levure, THO est de plus en plus recruté sur la PolII puis sur le transcrit au fur et à mesure que l'élongation progresse [100].

L'association de THO avec Yra1 et Sub2 est nécessaire à l'exportation de l'ARNm, puisque Yra1 est recruté après l'action de Sub2 et sert d'adaptateur pour le facteur d'exportation d'ARNm Mex67/Mtr2 chez la levure, NXF1/NXT1 (TAP•p15) chez l'homme. [101]. Le complexe Mex67/Mtr2 interagit avec le pore nucléaire et permet l'export de la mRNP.

Outre ce rôle clé dans l'exportation et le traitement de l'ARNm, THO joue un rôle important dans la prévention des R-loops. Les R-loops sont un phénomène courant se produisant lors de la transcription, où l'ADN de la bulle de transcription interagit avec l'ARN naissant, provoquant un échec de transcription et une instabilité génétique. Il n'est pas clair si les R-loops bloquent la PolII à partir de laquelle le fragment ARN est

né ou si les R-loops bloquent la PolII en amont [102]. Thp2 est au moins partiellement impliqué dans la résolution des R-loops, en particulier sur les locus [103].

Dans les cas où les mRNPs sont mal formées, notamment à cause d'un dysfonctionnement de THO, un système de contrôle qualité va prendre en charge les ARN malformés et empêcher leur export. La prochaine partie de ce résumé s'attardera sur ce système de contrôle qualité.

1.4 Contrôle qualité (QC) des mRNP

La machinerie de biogenèse du mRNP est robuste, mais pas infaillible. Pour atténuer d'éventuelles erreurs, les cellules ont développé un système de contrôle qualité des mRNPs. Ce système peut cibler et détruire les mRNPs formées de manière aberrante. Dans ce résumé, je ne parlerai que du système QC de la levure.

L'accumulation de transcrits aberrants dans les cellules est délétère de deux manières. Premièrement, les mRNPs aberrantes sont en compétition avec celles correctement formées pour divers facteurs protéiques, à la fois dans le noyau et dans le cytoplasme. Deuxièmement, la traduction d'un transcrit aberrant pourrait conduire à la formation d'une protéine délétère. Par conséquent, la dégradation de ces produits de transcriptions défectueuses est critique pour la cellule. La dégradation de ces transcrits est assurée par l'exosome, un complexe composé de 12 protéines principales, dont 9 sont organisées en structure en forme de tonneau, appelée ci-dessous le cœur ou le noyau de l'exosome. De plus, 2 exonucléases sont positionnées à chaque extrémité du cœur, Dis3 en bas et Rrp6 en haut. Une hélicase supplémentaire, Mtr4, est positionnée au-dessus de Rrp6 et déstructure le transcrit à dégrader. L'exosome a plusieurs types de cibles, gérés soit par Rrp6, Dis3 ou les deux [111]. En plus de leurs cibles qui se chevauchent, Rrp6 est connu pour stimuler allostériquement l'activité de Dis3 [112].

Comme dit précédemment, le noyau principal de l'exosome est composé de 9 protéines formant une structure en forme de tonneau à deux anneaux, communément appelée Exo9. L'anneau supérieur est composé de 3 sous-unités (Rrp4, Rrp40 et Csl4) appelé « cap ». Ils possèdent des domaines S1/KH similaires à ceux trouvés dans les protéines liant l'ARN. Pour former l'holocomplexe qu'est l'exosome (Exo14), le noyau de l'exosome est complété par 5 protéines, 1 située sous la forme du tonneau et les

autres sur sa face supérieure. La protéine située sous le noyau est Rrp44 ou Dis3, l'une des deux exonucléases du complexe. Les 4 autres protéines sont Rrp6, la deuxième exonucléase, avec ses deux cofacteurs, Rrp47, nécessaire à la stabilité de Rrp6 [113, 114], et Mpp6 [115]. La dernière protéine est Mtr4, l'hélicase associée au complexe. L'association de ces deux dernières protéines dépend de la présence à la fois de Rrp6 et de Rrp47 [116]. Des études récentes menées in vitro et in vivo suggèrent que Mpp6 est un autre acteur capable de lier Mtr4 à l'exosome chez l'homme [117, 118].

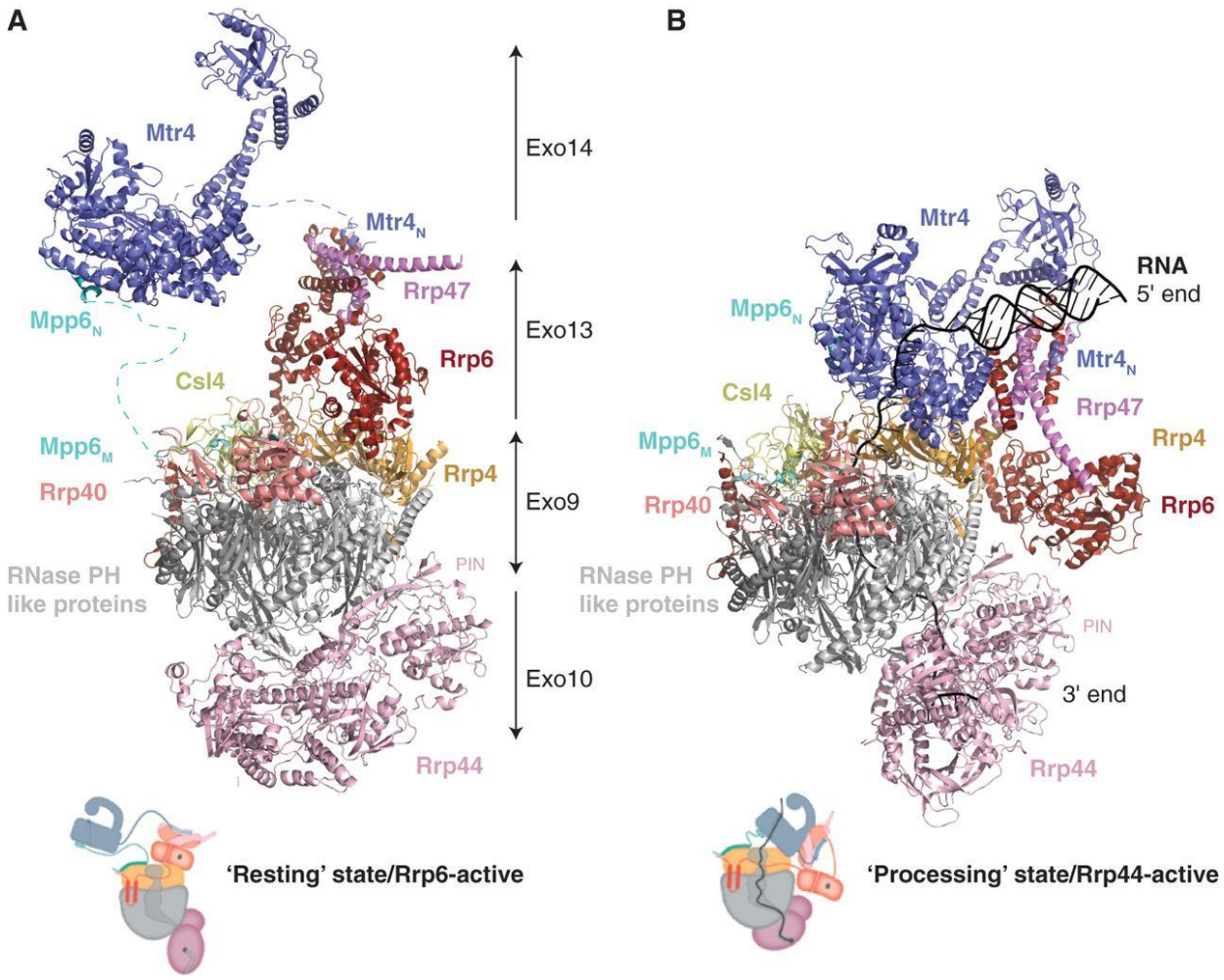


Figure R2 : Architecture de l'exosome. Ces modèles sont des superpositions de plusieurs structures cristallines. Exo9-10-13-14 sont les noms utilisés pour définir l'exosome en fonction des sous-unités qui lui sont associées. (A) Structure composite de l'exosome eucaryote au repos. Le noyau hélicase de Mtr4 est représenté dans un arrangement hypothétique. Les séquences carboxy -terminales de faible complexité de Mpp6 et Rrp6 sont flexibles et non représentées. (B) Structure composite de l'exosome eucaryote en cours de process. Tiré de [111]

Le cœur est traversé par un canal central qui servira au guidage de l'ARN cible vers l'exonucléase appropriée. L'activité hélicase de Mtr4 est nécessaire pour déstructurer l'ARN et l'introduire dans le canal central de l'exosome. Cependant, seul Dis3 traite les transcrits passant par ce canal tandis que les transcrits accèdent à Rrp6 en passant par l'anneau de la « cap » S1/KH [119]. Les ARN peuvent également accéder directement à Dis3 sans passer par le canal central de l'exosome [120].

Comme mentionné ci-dessus, Rrp6 est l'une des deux exonucléases 3'-5' de l'exosome. Elle est impliqué dans le traitement et la dégradation de l'ARN, ce qui comprend celle des ARN aberrants. Rrp6 est présent dans le noyau chez la levure et chez l'homme (connu sous le nom d'Exosc10 chez ce dernier), bien que Rrp6 ait également été détecté dans le cytoplasme uniquement chez l'homme [127, 128].

Bien que Rrp6 ne soit pas critique pour la survie des levures, elle est nécessaire pour le clivage de l'extrémité 3' du pré- sn/snoARN, la dégradation des CUT et des ARNm aberrants et la régulation de la longueur de la queue polyA [129-131]. Rrp6 est également responsable de la maturation des ARNr 5.8S et, aux côtés de TRAMP5 (Trf5-Air1/2-Mtr4), contribue à la dégradation des pré-ARNr aberrants. [132]. Certaines études ont montré que Dis3 et Rrp6 ont des rôles à la fois communs et distincts dans la dégradation et/ou le traitement de diverses classes d'ARN [133]. Notamment, Rrp6 est lié au traitement du 3' d'ARNnc spécifiques, qui ne sont pas ciblés par le mécanisme canonique de terminaison dépendant de polyA. [134]

De plus, certaines preuves suggèrent que Rrp6 et Dis3 peuvent conserver certains effets même lorsqu'ils ne sont pas liés au noyau de l'exosome [135-137]. La délétion de Rrp6, Dis3 et Rrp43 a montré que chacun d'eux était crucial pour la dégradation de classes spécifiques d'ARN [138]. Une étude récente a attribué à Rrp6 un rôle de régulateur de la dégradation des ARNt par désadénylation des transcrits traités par TRAMP [125].

Si l'exosome assure l'intégralité du processus catalytique de dégradation de l'ARN, d'autres acteurs sont nécessaires pour cibler ces transcrits. Ces acteurs sont intimement liés à la formation du mRNP car ces mêmes acteurs sont responsables à la fois de la protection ou du traitement et du ciblage. Par exemple, Nrd1 est une protéine clé dans le traitement du snoRNA et du ncRNA, mais il a également été

démontré qu'elle joue un rôle dans le ciblage et la dégradation des transcrits aberrants au sein du complexe NNS ainsi que du complexe TRAMP dans des conditions de stress. La structure et le fonctionnement de ces cofacteurs sont détaillés dans le manuscrit complet.

1.5 Le facteur Rho comme modèle expérimental

L'étude des mécanismes de contrôle qualité est ardue car le nombre de mRNP aberrantes générées dans des conditions physiologiques est très faible. Pour augmenter ce nombre, nous avons utilisé un facteur nommé Rho pour induire la formation de mRNP aberrants sans influencer le système QC, en évitant de supprimer un ou plusieurs de ses composants, comme cela a été fait dans des travaux antérieurs. Nous avons couplé l'induction Rho avec du ChIP-seq, nous permettant d'enregistrer des événements se produisant à l'échelle du génome. Dans cette partie, je traiterai du facteur Rho, utilisé pour nos expériences.

Découvert en 1969, Rho est un complexe homohexamérique couramment trouvé dans les bactéries [198]. Chez les procaryotes, il est responsable d'environ 20 à 30 % des événements de terminaison de la transcription chez les bactéries [199] et plus spécifiquement d'environ 50 % chez E.coli [200]. Outre ce rôle principal, Rho supprime également la transcription antisens [201], régule la population de transcrits lorsque la traduction est altérée [202]. Enfin, il résout les conflits entre les machineries de transcription et de réplication [203].

La manière dont Rho pourrait déclencher l'arrêt de la transcription est également débattue [200]. Les contraintes imposées par son activité hélicase aux brins hybrides ARN/ADN au sein de la bulle de transcription pourraient favoriser la terminaison de la transcription [211]. D'un autre côté, la force brute imposée par l'activité translocase de Rho pourrait être suffisante pour perturber l'hybride ADN/ARN au sein de la bulle de transcription, tirant l'ARN naissant hors de l'ARN polymérase et de la bulle de transcription [212]. Cette hypothèse est renforcée par le fait que Rho est capable de dissocier la streptavidine de l'ARN biotinylé [204]. En plus de sa capacité à désengager la polymérase de sa matrice ADN, Rho est connu pour être capable de rentrer en compétition avec d'autres protéines pour la liaison de l'ARN, une

caractéristique utilisée comme mécanisme de régulation de la transcription chez les procaryotes [213].

Comme mentionné ci-dessus, Rho est capable de perturber les fortes interactions ARN-protéines. Un grand nombre de protéines étant impliquées dans la formation de mRNPs chez les eucaryotes, l'équipe de Rachid Rahmouni a cherché à utiliser cette propriété de Rho comme inducteur de biogenèse aberrante d'ARNm. Par la suite, un modèle a été développé dans lequel Rho était induit chez *S. cerevisiae* sous un promoteur TetO7, réprimé par la doxycycline [25, 113, 176, 215-217]. Le même modèle a été utilisé pour créer une application directe de l'effet d'induction de la biogenèse de mRNPs aberrant de Rho, dans lequel il est utilisé comme outil de screening pour la découverte de nouveaux médicaments ciblant les bactéries [218].

La dernière étude de l'équipe a démontré que Rho possède une grande variété de cibles. En utilisant les techniques NGS, cette étude a montré qu'environ un transcrit sur cinq était affecté par Rho dans la levure. Un autre effet de l'induction de Rho chez la levure est un défaut de croissance associé à la dégradation des mRNPs aberrantes induites par Rho. Ce facteur a été utilisé pour identifier les protéines impliquées dans le ciblage et la dégradation des mRNPs aberrantes. [113, 176, 216]. Récemment, nous avons montré qu'en cas de perturbation de la biogenèse des mRNPs, certains facteurs responsables du traitement des ARNnc dans des conditions normales sont déplacés vers des locus codant pour l'ARNm [25].

À l'aide de cet outil, l'équipe a également mis en évidence une possible voie de dégradation parallèle des mRNPs aberrant induit par Rho, médiée par l'exonucléase 5'-3' Rat1, impliquée dans les modèles de terminaison de la transcription actuellement proposés par la communauté scientifique [216].

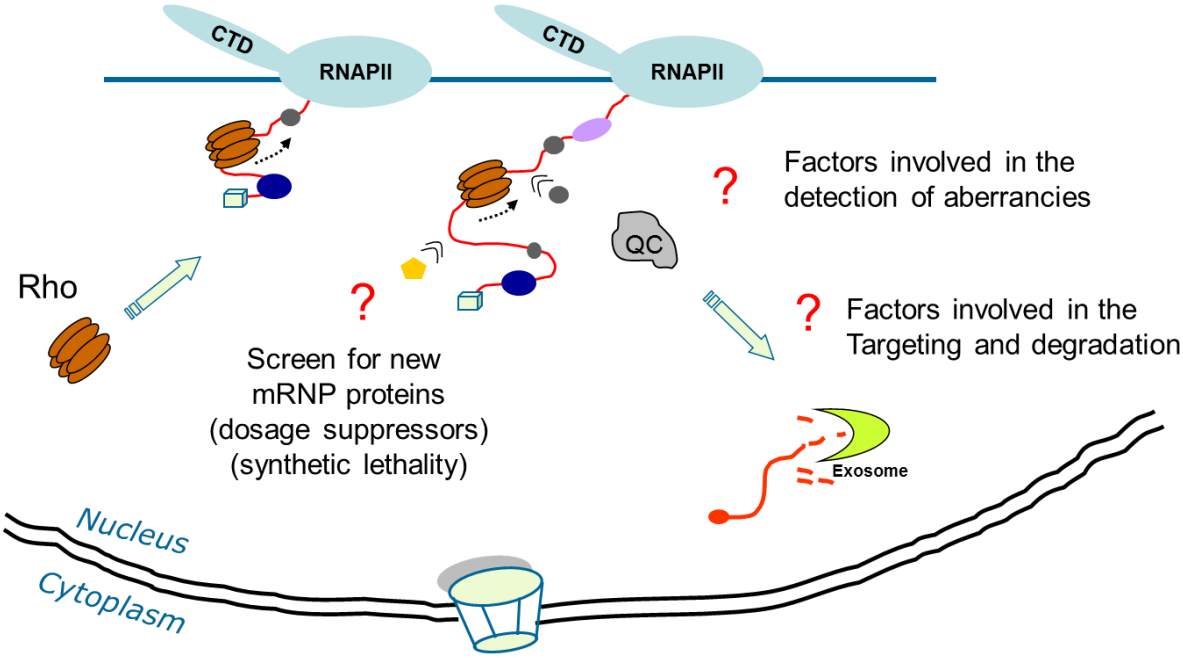


Figure R3 : Un modèle de la fonction de Rho.

1.6 Objectifs

Le fonctionnement de la machinerie de dégradation est bien documenté. Cependant, la détection et le ciblage des transcriptions aberrantes restent flous. La difficulté d'étudier ce mécanisme de ciblage est dû à deux facteurs. Premièrement, le contrôle qualité des mRNPs englobe une variété de complexes protéiques qui peuvent interagir les uns avec les autres directement ou indirectement via le transcrit d'ARN ou via le PolII. Ces interactions dépendent principalement du gène transcrit, comme précédemment identifié par notre équipe [25]. Ainsi, une étude axée sur un seul gène pourrait compromettre notre compréhension de l'ensemble du mécanisme de ciblage. Deuxièmement, la PolII et l'assemblage des mRNPs qui en résulte sont des systèmes résilients qui sont peu susceptibles de produire des mRNPs aberrants. Cela laisse une fenêtre d'opportunité très étroite pour étudier la dégradation des mRNPs aberrantes, un problème que nous sommes actuellement technologiquement incapables de résoudre, même avec les NGS. Par conséquent, l'utilisation d'un modèle comme Rho, conduisant à la génération de transcrits aberrantes, est presque inévitable.

Mon sujet trouve son origine dans des travaux antérieurs de mon équipe. Ayant utilisé Rho depuis plus d'une décennie, nous savons que Rho induit divers phénomènes physiologiques chez la levure, tels qu'une croissance plus lente et un recrutement élevé de Rrp6 sur les loci activement transcrits dans tout le génome. Dans

la littérature, les mêmes effets ont été observés chez les mutants du complexe THO/Sub2. Nous avons émis l'hypothèse que Rho pourrait avoir un effet sur THO. Ainsi, nous avons étudié l'effet de Rho sur le recrutement de THO sur des transcrits aberrants, pour voir si THO ou l'une de ses sous-unités jouait un rôle dans le ciblage des mRNPs aberrantes.

En parallèle, en collaboration avec l'équipe du Dr Igor Stuparevic à Zagreb, nous avons montré que Rrp6 régule l'expression de nombreux gènes codant pour des protéines cruciales de la paroi cellulaire. Le but de notre équipe dans ce projet était de confirmer les effets de la suppression de Rrp6 par analyse métagénomique de l'expression de gènes liés à la paroi cellulaire en fonction de leur ontologie. Nous avons également surveillé les effets de cette suppression sur l'expression d'ARNnc antisens ou promoteur-proximaux.

2. Résultats

Dans cette partie, je vais résumer brièvement les résultats obtenus durant mon doctorat. Du contenu supplémentaire est disponible dans la version complète du manuscrit. Ces résultats ont été soumis dans des journaux scientifiques et les manuscrits sont également accessibles en entier dans la partie résultats en anglais.

Article 1: Tho2 moonlights in yeast co-transcriptional mRNP quality control by targeting aberrant mRNPs to Rrp6

Dans cet article, nous montrons dans un premier temps que le recrutement de THO est perturbé lorsque Rho est induit. Plus particulièrement, nous montrons que les sous unités de THO, sauf une, semblent moins recrutées sur la chromatine. Contrairement aux autres, la sous unité Tho2 ne voit pas son recrutement baisser, mais augmenter lorsque Rho est induit. En traçant les métaprofils de recrutement de ces protéines sur la chromatine, nous avons pu à la fois vérifier que le recrutement de THO sans Rho était similaire à ce qui avait été documenté dans la littérature mais également voir que le recrutement de Tho2 était très différent d'un recrutement typique de THO lorsque Rho était induit. Dans un second temps, nous montrons que le sur-recrutement de Tho2 et la perte de recrutement des autres sous unités semble se corréler avec l'augmentation du recrutement de Rrp6 (une exonucléase de l'exosome) lorsque Rho

est induit. Pour finir nous montrons que ce recrutement de Rrp6 est Tho2 dépendant. Lorsque Tho2 est absent ou qu'il n'a plus la faculté d'interagir avec les acides nucléiques, il n'est plus recruté lorsque Rho est induit, tout comme Rrp6. De plus, des analyses sur le positionnement de ces deux protéines suggèrent que Tho2 et Rrp6 semblent être plus proche lorsque Rho est induit.

Article 2 : Yeast RNA exosome activity is necessary for maintaining cell wall stability through proper protein glycosylation

Dans ce second article, nous avons étudié l'impact de la délétion de Rrp6 sur le phénotype de levure *Saccharomyces Cerevisiae* avec nos collaborateurs de l'université de Zagreb. Pour rentrer dans les détails, les levures dépourvues de Rrp6 sont vulnérables à la chaleur et nous avons souhaité comprendre ce phénomène. Dans un premier temps, nous avons observé que l'ajout d'un stabilisateur osmotique permettaient de pallier cette vulnérabilité à la chaleur, ce qui pouvait signifier que la délétion de Rrp6 impactait la paroi cellulaire des levures. Nous avons ensuite ré-analysé des données de RNA-seq mesurant l'expression des gènes dans des souches sauvages ou sans Rrp6 à basse, moyenne et haute température. Grâce à cette approche, nous avons identifié plusieurs gènes candidats qui étaient moins exprimés à haute température dans les souches dépourvues de Rrp6 par rapport au sauvage. Nous avons remarqué qu'une surexpression de ces gènes à hautes températures dans les mêmes souches réduisait leur sensibilité à la chaleur. Pour finir, nous avons voulu lier Rrp6 et la baisse d'expression de certains gènes. Nous avons remarqué que certains des gènes affectés possédaient un CUT (long ARN non codant) à proximité de leur promoteur ou en antisens. Or, l'expression de ces CUT est modulée par Rrp6. Nous en avons déduit un modèle selon lequel la perte de Rrp6 provoque une transcription incontrôlée des CUT, qui perturbe la transcription de gènes codants à proximité, dont certains sont impliqués dans la synthèse de la paroi cellulaire des levures.

3. Discussion

3.1. Tho2 moonlights in yeast co-transcriptional mRNP quality control by targeting aberrant mRNPs to Rrp6

Grâce à la démocratisation des techniques NGS et plus globalement du domaine de recherche « omics », nous avons pu évaluer le fonctionnement du système de dégradation QC dans son ensemble. En surveillant le recrutement de composants clés du système aberrant de ciblage et de dégradation des mRNP, nous avons souligné l'importance de Tho2, une protéine impliquée dans l'empaquetage des mRNP. Dans le système canonique, Tho2 s'associe à 3 autres protéines pour former le complexe THO. Ce complexe est responsable du conditionnement précoce des mRNP et est nécessaire à leur exportation. L'interaction entre Tho2 et les mRNP a des résultats différents si Tho2 est impliqué ou non dans le complexe THO. Lorsque Tho2 interagit seul avec le transcrit, cela déclenche le recrutement du système de dégradation, c'est-à-dire l'exosome. On ne sait pas si le recrutement est direct ou dépend de cofacteurs, car d'autres facteurs sont également capables de recruter l'exosome, comme Isw1 [251]. De plus, Gbp2, une protéine déjà connue pour être impliquée dans l'épissage et le contrôle qualité, est capable de relier Tho2 et Rrp6 [252]. On peut supposer que Tho2 et Gbp2 peuvent interagir lorsque les deux sont sous forme monomérique. Le site d'interaction de Tho2 avec Gbp2 pourrait être masqué par d'autres sous-unités de THO lorsque Tho2 est dans le complexe. Seul, ce site pourrait être accessible à Gbp2 et l'interaction Tho2-Gbp2 pourrait conduire directement ou indirectement au recrutement de Rrp6/exosome. L'implication de complexes dans la dégradation de l'ARNm, normalement impliqués dans d'autres voies de transcription et/ou de dégradation, n'est pas nouveau. Par exemple, bien que le complexe NNS soit un élément clé dans le traitement et la dégradation de certains ARNnc, il joue également un rôle dans la dégradation des mRNP aberrants. [25, 253].

Il est clair que notre contribution ne conclut pas l'histoire de la relation entre Tho2 et Rrp6. Hormis un éventuel lien physique entre Rrp6 et Tho2, qui reste à démontrer, de nombreuses protéines pourraient constituer un lien entre Rrp6 et Tho2. En effet, tout cofacteur de Rrp6 ou Tho2 pourrait être considéré comme un candidat potentiel. L'étude de ces candidats potentiels et des interactions Rrp6 – Tho2 est un projet complet et ambitieux qui pourrait utiliser des techniques puissantes telles que

l'IP-MS (spectrométrie de masse par immunoprécipitation) pour le criblage suivi de Co-IP des candidats sélectionnés. Des techniques combinant ChIP et MS pourraient également être utilisées pour déterminer les interactions fonctionnelles de Tho2, Rrp6 et de leurs éventuels cofacteurs avec la chromatine. Dans ce cas, les locus enrichis en Tho2 lors de l'induction de Rho identifiés dans cette étude pourraient être réutilisés pour valider l'expérience. Un moyen plus intéressant et plus rapide d'effectuer une telle quantité d'analyses serait d'utiliser le co-fractionnement/spectrométrie de masse (CF/MS, [254]) sans induction de Rho pour répertorier tous les complexes Rrp6-Tho2 et énumérer leurs cofacteurs.

3.2. Yeast RNA exosome activity is necessary for maintaining cell wall stability through proper protein glycosylation.

La combinaison de l'approche bioinformatique avec un aspect expérimental plus traditionnel nous a permis d'explorer plus en profondeur certains mécanismes de l'exosome et de révéler les conséquences de sa déplétion sur la régulation de la transcription. On sait que la suppression de Rrp6 conduit à la régulation négative d'un grand nombre de gènes [25]. La délétion de Rrp6 chez la levure entraîne un défaut de croissance et une sensibilité à la température. Cette sensibilité à la température est réduite par l'ajout d'un stabilisant osmotique au milieu de culture. Des analyses bioinformatiques ont permis d'identifier des gènes impliqués dans l'intégrité de la paroi cellulaire. L'expression de ces gènes a ensuite été quantifiée par qPCR pour valider les résultats. En analysant l'expression des ARNnc en antisens et à proximité des promoteurs, nous avons constaté qu'un gène particulier (PSA1), qui est fortement régulé négativement lors de la suppression de Rrp6, avait son promoteur recouvert par un CUT hautement exprimé dans les mêmes conditions. Notre travail montre l'importance de la transcription pervasive pour la régulation de l'expression des gènes, dans la même veine que d'autres travaux récents [22, 256].

4. Conclusion

Au cours de mon doctorat, j'ai exploré le mécanisme de ciblage des mRNPs aberrantes par l’exosome en utilisant Rho comme perturbateur de la biogenèse des mRNPs. Mes travaux ont bénéficié de l’utilisation de Rho comme modèle pour étudier le contrôle qualité des mRNPs chez la levure, initiée par l’équipe du Dr Rachid Rahmouni il y a

un peu plus d'une décennie [217]. Des travaux récents élargissent notre spectre de recherche en utilisant la bioinformatique pour évaluer l'effet du modèle Rho sur l'ensemble de la population de transcrits [25]. Dans la continuité de l'approche développée par le Dr Kévin Moreau, j'ai adopté la bioinformatique pour étudier le complexe THO. Comme la plupart des travaux scientifiques, mes travaux sur le complexe THO ont soulevé autant de questions qu'ils en ont répondu. Nous avons montré que Tho2 et Rrp6 sont effectivement liés, au moins fonctionnellement. Cependant, on ne sait toujours pas comment et par quelles protéines cette connexion est établie, si ce n'est pas par Rrp6 et Tho2 eux-mêmes. L'ensemble du système qui contrôle la dégradation aberrante de l'ARNm doit continuer à être exploré.

Compte tenu des résultats obtenus avec la souche *Δrrp6* et d'autres travaux sur les défauts de transcription des ARNnc, nous avons appris que le manque de facteurs de dégradation des ARNnc conduit à des erreurs de terminaison, et il a été démontré que cela pourrait entraver la transcription des gènes codant pour l'ARNm en aval du gène produisant un ARNnc. Ce phénomène est connu sous le nom d'interférence transcriptionnelle [22]. Nous avons montré que l'effet global du défaut de terminaison de l'ARNnc et la perturbation de la transcription des gènes codants par la suppression de Rrp6 entraînent un défaut de croissance chez la levure dû à cette interférence transcriptionnelle.

5. References

1. Weiss, S.B. and L. Gladstone, *A MAMMALIAN SYSTEM FOR THE INCORPORATION OF CYTIDINE TRIPHOSPHATE INTO RIBONUCLEIC ACID1.* Journal of the American Chemical Society, 1959. **81**: p. 4118-4119.

2. Roeder, R.G. and W.J. Rutter, *Multiple Forms of DNA-dependent RNA Polymerase in Eukaryotic Organisms.* Nature, 1969. **224**(5216): p. 234-237.

3. Grummt, I., *Regulation of Mammalian Ribosomal Gene Transcription by RNA Polymerase I*, in *Progress in Nucleic Acid Research and Molecular Biology*, K. Moldave, Editor. 1998, Academic Press. p. 109-154.

4. Lee, Y., et al., *MicroRNA genes are transcribed by RNA polymerase II.* The EMBO Journal, 2004. **23**(20): p. 4051-4060.

5. WILLIS, I.M., *RNA polymerase III.* European Journal of Biochemistry, 1993. **212**(1): p. 1-11.

6. Griswold, A., *Genome packaging in prokaryotes: the circular chromosome of E. coli.* Nature Education, 2008. **1**(1): p. 57.

7. Lopes, I., et al., *Gene Size Matters: An Analysis of Gene Length in the Human Genome.* Frontiers in Genetics, 2021. **12**.

8. Tennyson, C.N., H.J. Klamut, and R.G. Worton, *The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced.* Nat Genet, 1995. **9**(2): p. 184-90.

9. Mignone, F., et al., *Untranslated regions of mRNAs.* Genome Biology, 2002. **3**(3): p. reviews0004.1.

10. Pesole, G., et al., *Structural and functional features of eukaryotic mRNA untranslated regions.* Gene, 2001. **276**(1-2): p. 73-81.

11. Hughes, M.J. and D.W. Andrews, *A single nucleotide is a sufficient 5' untranslated region for translation in an eukaryotic in vitro system.* FEBS Lett, 1997. **414**(1): p. 19-22.

12. van der Velden, A.W. and A.A. Thomas, *The role of the 5' untranslated region of an mRNA in translation regulation during development.* Int J Biochem Cell Biol, 1999. **31**(1): p. 87-106.

13. Jansen, R.P., *mRNA localization: message on the move.* Nat Rev Mol Cell Biol, 2001. **2**(4): p. 247-56.

14. Bashirullah, A., R.L. Cooperstock, and H.D. Lipshitz, *Spatial and temporal control of RNA stability.* Proc Natl Acad Sci U S A, 2001. **98**(13): p. 7025-8.

15. Conne, B., A. Stutz, and J.D. Vassalli, *The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology?* Nat Med, 2000. **6**(6): p. 637-41.

16. Brogna, S. and J. Wen, *Nonsense-mediated mRNA decay (NMD) mechanisms.* Nature Structural & Molecular Biology, 2009. **16**(2): p. 107-113.

17. Toma, K.G., et al., *Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay.* Rna, 2015. **21**(5): p. 887-97.

18. Sweeney, R., Q. Fan, and M.C. Yao, *Antisense ribosomes: rRNA as a vehicle for antisense RNAs.* Proc Natl Acad Sci U S A, 1996. **93**(16): p. 8518-23.

19. Beisang, D., et al., *Regulation of CUG-binding protein 1 (CUGBP1) binding to target transcripts upon T cell activation.* J Biol Chem, 2012. **287**(2): p. 950-60.

20. Wei, L.-H. and J.U. Guo, *Coding functions of "noncoding" RNAs.* Science, 2020. **367**(6482): p. 1074-1075.

21. Schmid, M. and T.H. Jensen, *Nuclear quality control of RNA polymerase II transcripts.* WIREs RNA, 2010. **1**(3): p. 474-485.

22. Villa, T. and O. Porrua, *Pervasive transcription: a controlled risk.* The FEBS Journal, 2022. **n/a**(n/a).

23. Barman, P., D. Reddy, and S.R. Bhaumik, *Mechanisms of Antisense Transcription Initiation with Implications in Gene Expression, Genomic Integrity and Disease Pathogenesis.* Noncoding RNA, 2019. **5**(1).

24. Patel, H.P., et al., *DNA supercoiling restricts the transcriptional bursting of neighboring eukaryotic genes.* bioRxiv, 2022: p. 2022.03.04.482969.

25. Moreau, K., et al., *Perturbation of mRNP biogenesis reveals a dynamic landscape of the Rrp6-dependent surveillance machinery trafficking along the yeast genome.* RNA Biology, 2019. **16**(7): p. 879-889.

26. van Dijk, E.L., et al., *XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast.* Nature, 2011. **475**(7354): p. 114-7.

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

28. Krahn, N., J.T. Fischer, and D. Söll, *Naturally Occurring tRNAs With Non-canonical Structures.* Frontiers in Microbiology, 2020. **11**.

29. Suzuki, T., *The expanding world of tRNA modifications and their disease relevance.* Nat Rev Mol Cell Biol, 2021. **22**(6): p. 375-392.

30. Sloan, K.E., et al., *Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function.* RNA Biology, 2017. **14**(9): p. 1138-1152.

31. Decatur, W.A. and M.J. Fournier, *rRNA modifications and ribosome function.* Trends Biochem Sci, 2002. **27**(7): p. 344-51.

32. Thomson, E. and D. Tollervey, *The final step in 5.8S rRNA processing is cytoplasmic in Saccharomyces cerevisiae.* Mol Cell Biol. **30**(4): p. 976-84.

33. Dieci, G., M. Preti, and B. Montanini, *Eukaryotic snoRNAs: a paradigm for gene expression flexibility.* Genomics, 2009. **94**(2): p. 83-8.

34. Huang, Z.-h., et al., *snoRNAs: functions and mechanisms in biological processes, and roles in tumor pathophysiology.* Cell Death Discovery, 2022. **8**(1): p. 259.

35. Williams, G.T. and F. Farzaneh, *Are snoRNAs and snoRNA host genes new players in cancer?* Nat Rev Cancer, 2012. **12**(2): p. 84-8.

36. Czekay, D.P. and U. Kothe, *H/ACA Small Ribonucleoproteins: Structural and Functional Comparison Between Archaea and Eukaryotes.* Frontiers in Microbiology, 2021. **12**.

37. Gumienny, R., et al., *High-throughput identification of C/D box snoRNA targets with CLIP and RiboMeth-seq.* Nucleic Acids Res, 2017. **45**(5): p. 2341-2353.

38. van Hoof, A., P. Lennertz, and R. Parker, *Yeast exosome mutants accumulate 3'-extended polyadenylated forms of U4 small nuclear RNA and small nucleolar RNAs.* Mol Cell Biol, 2000. **20**(2): p. 441-52.

39. Jurica, M.S. and M.J. Moore, *Pre-mRNA splicing: awash in a sea of proteins.* Mol Cell, 2003. **12**(1): p. 5-14.

40. Morais, P., H. Adachi, and Y.-T. Yu, *Spliceosomal snRNA Epitranscriptomics.* Frontiers in Genetics, 2021. **12**.

41. Dana, H., et al., *Molecular Mechanisms and Biological Functions of siRNA.* Int J Biomed Sci, 2017. **13**(2): p. 48-57.

42. Drinnenberg, I.A., et al., *RNAi in budding yeast.* Science, 2009. **326**(5952): p. 544-550.

43. O'Brien, J., et al., *Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation.* Front Endocrinol (Lausanne), 2018. **9**: p. 402.

44. Friedrich, M. and A. Aigner, *Therapeutic siRNA: State-of-the-Art and Future Perspectives.* BioDrugs, 2022. **36**(5): p. 549-571.

45. Xu, W., et al., *Identifying microRNA targets in different gene regions.* BMC Bioinformatics, 2014. **15**(7): p. S4.

46. Dharap, A., et al., *MicroRNA miR-324-3p Induces Promoter-Mediated Expression of RelA Gene.* PLOS ONE, 2013. **8**(11): p. e79467.

47. Bhattacharjee, S., B. Roche, and R.A. Martienssen, *RNA-induced initiation of transcriptional silencing (RITS) complex structure and function.* RNA Biol, 2019. **16**(9): p. 1133-1146.

48. Creamer, K.M. and J.F. Partridge, *RITS-connecting transcription, RNA interference, and heterochromatin assembly in fission yeast.* Wiley Interdiscip Rev RNA, 2011. **2**(5): p. 632-46.

49. Ozata, D.M., et al., *PIWI-interacting RNAs: small RNAs with big functions.* Nature Reviews Genetics, 2019. **20**(2): p. 89-108.

50. Meyer, P.A., et al., *Structure of the 12-subunit RNA polymerase II refined with the aid of anomalous diffraction data.* J Biol Chem, 2009. **284**(19): p. 12933-9.

51. Schier, A.C. and D.J. Taatjes, *Structure and mechanism of the RNA polymerase II transcription machinery.* Genes Dev, 2020. **34**(7-8): p. 465-488.

52. He, Y., et al., *Near-atomic resolution visualization of human transcription promoter opening.* Nature, 2016. **533**(7603): p. 359-65.

53. Bernecky, C., et al., *Structure of transcribing mammalian RNA polymerase II.* Nature, 2016. **529**(7587): p. 551-4.

54. Zaborowska, J., S. Egloff, and S. Murphy, *The pol II CTD: new twists in the tail.* Nat Struct Mol Biol, 2016. **23**(9): p. 771-7.

55. Milligan, L., et al., *Strand-specific, high-resolution mapping of modified RNA polymerase II.* Molecular Systems Biology, 2016. **12**(6).

56. Spain, M.M. and C.K. Govind, *A role for phosphorylated Pol II CTD in modulating transcription coupled histone dynamics.* Transcription, 2011. **2**(2): p. 78-81.

57. Nojima, T., et al., *RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing.* Molecular Cell, 2018. **72**(2): p. 369-379.e4.

58. Guo, Y.E., et al., *Pol II phosphorylation regulates a switch between transcriptional and splicing condensates.* Nature, 2019. **572**(7770): p. 543-548.

59. Reines, D., *Recent advances in understanding RNA polymerase II structure and function.* Fac Rev, 2020. **9**: p. 11.

60. Zheng, Z.-L., *Cyclin-Dependent Kinases and CTD Phosphatases in Cell Cycle Transcriptional Control: Conservation across Eukaryotic Kingdoms and Uniqueness to Plants.* Cells, 2022. **11**(2): p. 279.

61. Donczew, R. and S. Hahn, *Mechanistic Differences in Transcription Initiation at TATA-Less and TATA-Containing Promoters.* Mol Cell Biol, 2018. **38**(1).

62. Roy, A.L. and D.S. Singer, *Core promoters in transcription: old problem, new insights.* Trends Biochem Sci, 2015. **40**(3): p. 165-71.

63. Sainsbury, S., C. Bernecky, and P. Cramer, *Structural basis of transcription initiation by RNA polymerase II.* Nature Reviews Molecular Cell Biology, 2015. **16**(3): p. 129-143.

64. Hahn, S., *Structure and mechanism of the RNA polymerase II transcription machinery.* Nat Struct Mol Biol, 2004. **11**(5): p. 394-403.

65. Core, L. and K. Adelman, *Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation.* Genes & Development, 2019.

66. Wong, K.H., Y. Jin, and K. Struhl, *TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape.* Mol Cell, 2014. **54**(4): p. 601-12.

67. Vos, S.M., et al., *Structure of paused transcription complex Pol II–DSIF–NELF.* Nature, 2018. **560**(7720): p. 601-606.

68. Jonkers, I. and J.T. Lis, *Getting up to speed with transcription elongation by RNA polymerase II.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 167-77.

69. Cheung, A.C.M. and P. Cramer, *Structural basis of RNA polymerase II backtracking, arrest and reactivation.* Nature, 2011. **471**(7337): p. 249-253.

70. Muniz, L., E. Nicolas, and D. Trouche, *RNA polymerase II speed: a key player in controlling and adapting transcriptome composition.* The EMBO Journal, 2021. **40**(15): p. e105740.

71. Lian, Z., et al., *A genomic analysis of RNA polymerase II modification and chromatin architecture related to 3' end RNA polyadenylation.* Genome Res, 2008. **18**(8): p. 1224-37.

72. Porrua, O. and D. Libri, *Transcription termination and the control of the transcriptome: why, where and how to stop.* Nat Rev Mol Cell Biol, 2015. **16**(3): p. 190-202.

73. Carminati, M., et al., *A direct interaction between CPF and Pol II links RNA 3'-end processing to transcription.* bioRxiv, 2022: p. 2022.07.28.501803.

74. Luo, W., A.W. Johnson, and D.L. Bentley, *The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model.* Genes Dev, 2006. **20**(8): p. 954-65.

75. Eaton, J.D., et al., *A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes.* Genes & Development, 2020. **34**(1-2): p. 132-145.

76. Schlackow, M., et al., *Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs.* Molecular Cell, 2016.

77. Eaton, J.D. and S. West, *Termination of Transcription by RNA Polymerase II: BOOM!* Trends Genet, 2020. **36**(9): p. 664-675.

78. Cortazar, M.A., et al., *Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism.* Mol Cell, 2019. **76**(6): p. 896-908.e4.

79. Creamer, T.J., et al., *Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1.* PLoS Genet, 2011. **7**(10): p. e1002329.

80. Tudek, A., et al., *Molecular basis for coordinating transcription termination with noncoding RNA degradation.* Mol Cell, 2014. **55**(3): p. 467-81.

81. Gonatopoulos-Pournatzis, T. and V.H. Cowling, *Cap-binding complex (CBC).* Biochem J, 2014. **457**(2): p. 231-42.

82. Isken, O. and L.E. Maquat, *Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function.* Genes Dev, 2007. **21**(15): p. 1833-56.

83. Mars, J.C., et al., *The Cap-Binding Complex CBC and the Eukaryotic Translation Factor eIF4E: Co-Conspirators in Cap-Dependent RNA Maturation and Translation.* Cancers (Basel), 2021. **13**(24).

84. Ryu, I. and Y.K. Kim, *Translation initiation mediated by nuclear cap-binding protein complex.* BMB Rep, 2017. **50**(4): p. 186-193.

85. Ares, M., Jr., L. Grate, and M.H. Pauling, *A handful of intron-containing genes produces the lion's share of yeast mRNA.* Rna, 1999. **5**(9): p. 1138-9.

86. Hossain, M.A. and T.L. Johnson, *Using yeast genetics to study splicing mechanisms.* Methods Mol Biol, 2014. **1126**: p. 285-98.

87. Wang, Y., et al., *Mechanism of alternative splicing and its regulation (Review).* Biomed Rep, 2015. **3**(2): p. 152-158.

88. Fair, B.J. and J.A. Pleiss, *The power of fission: yeast as a tool for understanding complex splicing.* Curr Genet, 2017. **63**(3): p. 375-380.

89. Gehring, N.H. and J.-Y. Roignant, *Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation.* Trends in Genetics, 2021. **37**(4): p. 355-372.

90. Chavez, S., et al., *A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae.* EMBO J, 2000. **19**(21): p. 5824-34.

91. Schuller, S.K., et al., *Structural insights into the nucleic acid remodeling mechanisms of the yeast THO-Sub2 complex.* eLife, 2020. **9**: p. e61467.

92. Masuda, S., et al., *Recruitment of the human TREX complex to mRNA during splicing.* Genes Dev, 2005. **19**(13): p. 1512-7.

93. Gwizdek, C., et al., *The mRNA nuclear export factor Hpr1 is regulated by Rsp5-mediated ubiquitylation.* J Biol Chem, 2005. **280**(14): p. 13401-5.

94. Xie, Y., et al., *Cryo-EM structure of the yeast TREX complex and coordination with the SR-like protein Gbp2.* bioRxiv, 2021: p. 2021.01.04.425184.

95. Pühringer, T., et al., *Structure of the human core transcription-export complex reveals a hub for multivalent interactions.* eLife, 2020. **9**: p. e61503.

96. Pena, A., et al., *Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor.* EMBO J, 2012. **31**(6): p. 1605-1616.

97. Hurt, E., et al., *Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex.* Proc Natl Acad Sci U S A, 2004. **101**(7): p. 1858-62.

98. Hackmann, A., et al., *Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1.* Nat Commun, 2014. **5**.

99. Meinel, D.M., et al., *Recruitment of TREX to the transcription machinery by its direct binding to the phospho-CTD of RNA polymerase II.* PLoS Genet, 2013. **9**(11): p. e1003914.

100. Gómez-González, B., et al., *Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles.* Embo j, 2011. **30**(15): p. 3106-19.

101. Xie, Y. and Y. Ren, *Mechanisms of nuclear mRNA export: A structural perspective.* Traffic, 2019. **20**(11): p. 829-840.

102. Luna, R., et al., *The THO Complex as a Paradigm for the Prevention of Cotranscriptional R-Loops.* Cold Spring Harb Symp Quant Biol, 2019. **84**: p. 105-114.

103. Pfeiffer, V., et al., *The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening.* The EMBO Journal, 2013. **32**(21): p. 2861-2871.

104. Rougemaille, M., et al., *THO/Sub2p functions to coordinate 3'-end processing with gene-nuclear pore association.* Cell, 2008. **135**(2): p. 308-21.

105. Rougemaille, M., et al., *Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants.* EMBO J, 2007. **26**(9): p. 2317-26.

106. Libri, D., et al., *Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation.* Mol Cell Biol, 2002. **22**(23): p. 8254-66.

107. Saguez, C., et al., *Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation.* Mol Cell, 2008. **31**(1): p. 91-103.

108. Tudek, A., M. Lloret-Llinares, and T.H. Jensen, *The multitasking polyA tail: nuclear RNA maturation, degradation and export.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2018. **373**(1762): p. 20180169.

109. Iglesias, N., et al., *Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export.* Genes Dev, 2010. **24**(17): p. 1927-38.

110. Stewart, M., *Polyadenylation and nuclear export of mRNAs.* Journal of Biological Chemistry, 2019. **294**(9): p. 2977-2987.

111. Lingaraju, M., et al., *To Process or to Decay: A Mechanistic View of the Nuclear RNA Exosome.* Cold Spring Harb Symp Quant Biol, 2019. **84**: p. 155-163.

112. Wasmuth, E.V. and C.D. Lima, *The Rrp6 C-terminal domain binds RNA and activates the nuclear RNA exosome.* Nucleic Acids Research, 2016.

113. Stuparevic, I., et al., *Cotranscriptional recruitment of RNA exosome cofactors Rrp47p and Mpp6p and two distinct Trf-Air-Mtr4 polyadenylation (TRAMP) complexes assists the exonuclease Rrp6p in the targeting and degradation of an aberrant messenger ribonucleoprotein particle (mRNP) in yeast.* J Biol Chem, 2013. **288**(44): p. 31816-29.

114. Feigenbutz, M., et al., *The Exosome Cofactor Rrp47 Is Critical for the Stability and Normal Expression of Its Associated Exoribonuclease Rrp6 in Saccharomyces cerevisiae.* PLOS ONE, 2013. **8**(11): p. e80752.

115. Falk, S., et al., *Mpp6 Incorporation in the Nuclear Exosome Contributes to RNA Channeling through the Mtr4 Helicase.* Cell Rep, 2017. **20**(10): p. 2279-2286.

116. Schuch, B., et al., *The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase.* The EMBO Journal, 2014. **33**(23): p. 2829-2846.

117. Fujiwara, N., et al., *MPP6 stimulates both RRP6 and DIS3 to degrade a specified subset of MTR4-sensitive substrates in the human nucleus.* Nucleic Acids Res, 2022. **50**(15): p. 8779-8806.

118. Wasmuth, E.V., et al., *Structure and reconstitution of yeast Mpp6-nuclear exosome complexes reveals that Mpp6 stimulates RNA decay and recruits the Mtr4 helicase.* eLife, 2017. **6**: p. e29062.

119. Zinder, J.C., E.V. Wasmuth, and C.D. Lima, *Nuclear RNA Exosome at 3.1 Å Reveals Substrate Specificities, RNA Paths, and Allosteric Inhibition of Rrp44/Dis3.* Mol Cell, 2016. **64**(4): p. 734-745.

120. Liu, J.J., et al., *Visualization of distinct substrate-recruitment pathways in the yeast exosome by EM.* Nat Struct Mol Biol, 2014. **21**(1): p. 95-102.

121. Han, J. and A. van Hoof, *The RNA Exosome Channeling and Direct Access Conformations Have Distinct In Vivo Functions.* Cell Rep, 2016. **16**(12): p. 3348-3358.

122. Schneider, C., et al., *Transcriptome-wide Analysis of Exosome Targets.* Mol Cell, 2012. **48**(3): p. 422-33.

123. Makino, D.L., M. Baumgärtner, and E. Conti, *Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex.* Nature, 2013. **495**(7439): p. 70-5.

124. Wasmuth, E.V. and C.D. Lima, *Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel.* Mol Cell, 2012. **48**(1): p. 133-44.

125. Das, M., et al., *Substrate discrimination and quality control require each catalytic activity of TRAMP and the nuclear RNA exosome.* Proceedings of the National Academy of Sciences, 2021. **118**(14): p. e2024846118.

126. Ogami, K., et al., *An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression.* Genes & Development, 2017. **31**(12): p. 1257-1271.

127. van Dijk, E.L., G. Schilders, and G.J.M. Pruijn, *Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways.* RNA, 2007. **13**(7): p. 1027-1035.

128. de Almeida, S.F., et al., *A link between nuclear RNA surveillance, the human exosome and RNA polymerase II transcriptional termination.* Nucleic Acids Res, 2010. **38**(22): p. 8015-26.

129. Vasiljeva, L. and S. Buratowski, *Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts.* Mol Cell, 2006. **21**(2): p. 239-48.

130. Colin, J., D. Libri, and O. Porrua, *Cryptic transcription and early termination in the control of gene expression.* Genet Res Int, 2011. **2011**: p. 653494.

131. Schmid, M. and T.H. Jensen, *Quality control of mRNP in the nucleus.* Chromosoma, 2008. **117**(5): p. 419-29.

132. Wery, M., et al., *The nuclear poly(A) polymerase and Exosome cofactor Trf5 is recruited cotranscriptionally to nucleolar surveillance.* Rna, 2009. **15**(3): p. 406-19.

133. Gudipati, R.K., et al., *Extensive degradation of RNA precursors by the exosome in wild-type cells.* Mol Cell, 2012. **48**(3): p. 409-21.

134. Thiebaut, M., et al., *Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance.* Mol Cell, 2006. **23**(6): p. 853-64.

135. Noguchi, E., et al., *Dis3, implicated in mitotic control, binds directly to Ran and enhances the GEF activity of RCC1.* The EMBO Journal, 1996. **15**(20): p. 5595-5605.

136. Graham, A.C., D.L. Kiss, and E.D. Andrulis, *Core Exosome-independent Roles for Rrp6 in Cell Cycle Progression.* Molecular Biology of the Cell, 2009. **20**(8): p. 2242-2253.

137. Wang, C., et al., *Rrp6 Moonlights in an RNA Exosome-Independent Manner to Promote Cell Survival and Gene Expression during Stress.* Cell Reports, 2020. **31**(10): p. 107754.

138. Callahan, K.P. and J.S. Butler, *Evidence for core exosome independent function of the nuclear exoribonuclease Rrp6p.* Nucleic Acids Res, 2008. **36**(21): p. 6645-55.

139. Fox, M.J. and A.L. Mosley, *Rrp6: Integrated roles in nuclear RNA metabolism and transcription termination.* Wiley Interdiscip Rev RNA, 2016. **7**(1): p. 91-104.

140. Schilders, G., et al., *MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation.* Nucleic Acids Res, 2005. **33**(21): p. 6795-804.

141. Milligan, L., et al., *A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts.* Mol Cell Biol, 2008. **28**(17): p. 5446-57.

142. Fasken, M.B., R.N. Laribee, and A.H. Corbett, *Nab3 Facilitates the Function of the TRAMP Complex in RNA Processing via Recruitment of Rrp6 Independent of Nrd1.* PLoS Genet, 2015. **11**(3): p. e1005044.

143. Heo, D.-h., et al., *The RNA Polymerase II C-terminal Domain-interacting Domain of Yeast Nrd1 Contributes to the Choice of Termination Pathway and Couples to RNA Processing by the Nuclear Exosome.* Journal of Biological Chemistry, 2013. **288**(51): p. 36676-36690.

144. Schneider, C., et al., *The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome.* Nucleic Acids Res, 2009. **37**(4): p. 1127-40.

145. Milligan, L., et al., *A nuclear surveillance pathway for mRNAs with defective polyadenylation.* Mol Cell Biol, 2005. **25**(22): p. 9996-10004.

146. Robinson, S.R., et al., *The 3' to 5' Exoribonuclease DIS3: From Structure and Mechanisms to Biological Functions and Role in Human Disease.* Biomolecules, 2015. **5**(3): p. 1515-39.

147. Allmang, C., et al., *Functions of the exosome in rRNA, snoRNA and snRNA synthesis.* EMBO J, 1999. **18**(19): p. 5399-410.

148. Mitchell, P., et al., *The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases.* Cell, 1997. **91**(4): p. 457-66.

149. Lorentzen, E., et al., *Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family.* Mol Cell, 2008. **29**(6): p. 717-28.

150. Zinder, J.C. and C.D. Lima, *Targeting RNA for processing or destruction by the eukaryotic RNA exosome and its cofactors.* Genes & Development, 2017. **31**(2): p. 88-100.

151. Kadaba, S., et al., *Nuclear surveillance and degradation of hypomodified initiator tRNAMet in S. cerevisiae.* Genes Dev, 2004. **18**(11): p. 1227-40.

152. LaCava, J., et al., *RNA degradation by the exosome is promoted by a nuclear polyadenylation complex.* Cell, 2005. **121**(5): p. 713-24.

153. Vanacova, S., et al., *A new yeast poly(A) polymerase complex involved in RNA quality control.* PLoS Biol, 2005. **3**(6): p. e189.

154. Callahan, K.P. and J.S. Butler, *TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6.* J Biol Chem, 2010. **285**(6): p. 3540-7.

155. Fasken, M.B., et al., *Air1 zinc knuckles 4 and 5 and a conserved IWRXY motif are critical for the function and integrity of the Trf4/5-Air1/2-Mtr4 polyadenylation (TRAMP) RNA quality control complex.* J Biol Chem, 2011. **286**(43): p. 37429-45.

156. Wyers, F., et al., *Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase.* Cell, 2005. **121**(5): p. 725-37.

157. Delan-Forino, C., et al., *Substrate specificity of the TRAMP nuclear surveillance complexes.* Nature Communications, 2020. **11**(1): p. 3122.

158. Wlotzka, W., et al., *The nuclear RNA polymerase II surveillance system targets polymerase III transcripts.* Embo j, 2011. **30**(9): p. 1790-803.

159. Dickinson, H., et al., *The TRAMP complex shows tRNA editing activity in S. cerevisiae.* Mol Biol Evol, 2012. **29**(5): p. 1451-9.

160. Kong, K.-Y.E., et al., *Cotranscriptional recruitment of yeast TRAMP complex to intronic sequences promotes optimal pre-mRNA splicing.* Nucleic Acids Research, 2014. **42**(1): p. 643-660.

161. Wong, C.-m., et al., *Current perspectives on the role of TRAMP in nuclear RNA surveillance and quality control.* Research and Reports in Biochemistry, 2015. **2015**: p. 111.

162. Ogami, K., Y. Chen, and J.L. Manley, *RNA Surveillance by the Nuclear RNA Exosome: Mechanisms and Significance.* Non-Coding RNA, 2018. **4**(1): p. 8.

163. Vanacova, S. and R. Stefl, *The exosome and RNA quality control in the nucleus.* EMBO Rep, 2007. **8**(7): p. 651-7.

164. Schulz, D., et al., *Transcriptome surveillance by selective termination of noncoding RNA synthesis.* Cell, 2013. **155**(5): p. 1075-87.

165. Arigo, J.T., et al., *Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3.* Mol Cell, 2006. **23**(6): p. 841-51.

166. Conrad, N.K., et al., *A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II.* Genetics, 2000. **154**(2): p. 557-71.

167. Steinmetz, E.J., et al., *RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts.* Nature, 2001. **413**(6853): p. 327-31.

168. Carroll, K.L., et al., *Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts.* Mol Cell Biol, 2004. **24**(14): p. 6241-52.

169. Jamonnak, N., et al., *Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing.* RNA, 2011. **17**(11): p. 2011-25.

170. Vasiljeva, L., et al., *The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain.* Nat Struct Mol Biol, 2008. **15**(8): p. 795-804.

171. Gudipati, R.K., et al., *Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice.* Nat Struct Mol Biol, 2008. **15**(8): p. 786-94.

172. Carroll, K.L., et al., *Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements.* RNA, 2007. **13**(3): p. 361-73.

173. Webb, S., et al., *PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast.* Genome Biology, 2014. **15**(1): p. R8.

174. Porrua, O., et al., *In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination.* EMBO J, 2012. **31**(19): p. 3935-48.

175. Larochelle, M., et al., *Common mechanism of transcription termination at coding and noncoding RNA genes in fission yeast.* Nat Commun, 2018. **9**(1): p. 4364.

176. Honorine, R., et al., *Nuclear mRNA quality control in yeast is mediated by Nrd1 co-transcriptional recruitment, as revealed by the targeting of Rho-induced aberrant transcripts.* Nucleic Acids Res, 2011. **39**(7): p. 2809-20.

177. Bresson, S. and D. Tollervey, *Surveillance-ready transcription: nuclear RNA decay as a default fate.* Open Biology, 2018. **8**(3): p. 170270.

178. Morton, D.J., et al., *The RNA exosome and RNA exosome-linked disease.* RNA, 2018. **24**(2): p. 127-142.

179. Liu, Q., J.C. Greimann, and C.D. Lima, *Reconstitution, activities, and structure of the eukaryotic RNA exosome.* Cell, 2006. **127**(6): p. 1223-37.

180. Makino, D.L., et al., *RNA degradation paths in a 12-subunit nuclear exosome complex.* Nature, 2015. **advance online publication**.

181. Tomecki, R., et al., *The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L.* Embo j, 2010. **29**(14): p. 2342-57.

182. Sloan, K.E., C. Schneider, and N.J. Watkins, *Comparison of the yeast and human nuclear exosome complexes.* Biochem Soc Trans, 2012. **40**(4): p. 850-5.

183. Gerlach, P., et al., *Distinct and evolutionary conserved structural features of the human nuclear exosome complex.* eLife, 2018. **7**: p. e38686.

184. Allmang, C., et al., *The yeast exosome and human PM–Scl are related complexes of 3′ → 5′ exonucleases.* Genes & Development, 1999. **13**(16): p. 2148-2158.

185. Preker, P., et al., *RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters.* Science, 2008. **322**(5909): p. 1851-1854.

186. Flynn, R.A., et al., *Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome.* Proc Natl Acad Sci U S A, 2011. **108**(26): p. 10460-5.

187. Pefanis, E., et al., *Noncoding RNA transcription targets AID to divergently transcribed loci in B cells.* Nature, 2014. **514**(7522): p. 389-93.

188. Szczepińska, T., et al., *DIS3 shapes the RNA polymerase II transcriptome in humans by degrading a variety of unwanted transcripts.* Genome Research, 2015. **25**(11): p. 1622-1633.

189. Schilders, G., E. van Dijk, and G.J.M. Pruijn, *C1D and hMtr4p associate with the human exosome subunit PM/Scl-100 and are involved in pre-rRNA processing.* Nucleic Acids Research, 2007. **35**(8): p. 2564-2572.

190. Schilders, G., et al., *C1D is a major autoantibody target in patients with the polymyositis–scleroderma overlap syndrome.* Arthritis & Rheumatism, 2007. **56**(7): p. 2449-2454.

191. Lubas, M., et al., *Interaction profiling identifies the human nuclear exosome targeting complex.* Mol Cell, 2011. **43**(4): p. 624-37.

192. Berndt, H., et al., *Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming.* Rna, 2012. **18**(5): p. 958-72.

193. Nag, A. and J.A. Steitz, *Tri-snRNP-associated proteins interact with subunits of the TRAMP and nuclear exosome complexes, linking RNA decay and pre-mRNA splicing.* RNA Biol, 2012. **9**(3): p. 334-42.

194. Meola, N., et al., *Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts.* Mol Cell, 2016. **64**(3): p. 520-533.

195. Silla, T., et al., *The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts.* Cell Reports, 2018. **23**(7): p. 2199-2210.

196. Wu, G., et al., *A Two-Layered Targeting Mechanism Underlies Nuclear RNA Sorting by the Human Exosome.* Cell Reports, 2020. **30**(7): p. 2387-2401.e5.

197. Andersen, P.R., et al., *The human cap-binding complex is functionally connected to the nuclear RNA exosome.* Nat Struct Mol Biol, 2013. **20**(12): p. 1367-76.

198. Roberts, J.W., *Termination factor for RNA synthesis.* Nature, 1969. **224**(5225): p. 1168-74.

199. Kriner, M.A., A. Sevostyanova, and E.A. Groisman, *Learning from the Leaders: Gene Regulation by the Transcription Termination Factor Rho.* Trends Biochem Sci, 2016. **41**(8): p. 690-699.

200. Mitra, P., et al., *Rho Protein: Roles and Mechanisms.* Annu Rev Microbiol, 2017. **71**: p. 687-709.

201. Peters, J.M., et al., *Rho and NusG suppress pervasive antisense transcription in Escherichia coli.* Genes & Development, 2012. **26**(23): p. 2621-2633.

202. Richardson, J.P., *Preventing the synthesis of unused transcripts by Rho factor.* Cell, 1991. **64**(6): p. 1047-9.

203. Washburn, R.S. and M.E. Gottesman, *Transcription termination maintains chromosome integrity.* Proceedings of the National Academy of Sciences, 2011. **108**(2): p. 792-797.

204. Schwartz, A., et al., *Transcription termination factor rho can displace streptavidin from biotinylated RNA.* J Biol Chem, 2007. **282**(43): p. 31469-76.

205. Sevostyanova, A. and E.A. Groisman, *An RNA motif advances transcription by preventing Rho-dependent termination.* Proceedings of the National Academy of Sciences, 2015. **112**(50): p. E6835-E6843.

206. Richardson, J.P., *Rho-dependent termination and ATPases in transcript termination.* Biochim Biophys Acta, 2002. **1577**(2): p. 251-260.

207. Steinmetz, E.J. and T. Platt, *Evidence supporting a tethered tracking model for helicase activity of Escherichia coli Rho factor.* Proceedings of the National Academy of Sciences, 1994. **91**(4): p. 1401-1405.

208. Kriner, M.A. and E.A. Groisman, *The Bacterial Transcription Termination Factor Rho Coordinates Mg(2+) Homeostasis with Translational Signals.* J Mol Biol, 2015. **427**(24): p. 3834-49.

209. Figueroa-Bossi, N., et al., *RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination.* Genes Dev, 2014. **28**(11): p. 1239-51.

210. Bossi, L., et al., *A role for Rho-dependent polarity in gene regulation by a noncoding small RNA.* Genes Dev, 2012. **26**(16): p. 1864-73.

211. Brennan, C.A., A.J. Dombroski, and T. Platt, *Transcription termination factor rho is an RNA-DNA helicase.* Cell, 1987. **48**(6): p. 945-52.

212. Dutta, D., J. Chalissery, and R. Sen, *Transcription Termination Factor Rho Prefers Catalytically Active Elongation Complexes for Releasing RNA*.* Journal of Biological Chemistry, 2008. **283**(29): p. 20243-20251.

213. Vieu, E. and A.R. Rahmouni, *Dual role of boxB RNA motif in the mechanisms of termination/antitermination at the lambda tR1 terminator revealed in vivo.* J Mol Biol, 2004. **339**(5): p. 1077-87.

214. Skordalakes, E. and J.M. Berger, *Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading.* Cell, 2003. **114**(1): p. 135-46.

215. Mosrin-Huaman, C., Hervouet-Coste, N., Le Dantec, A., Stuparevic, I., Rahmouni, A.R., *Bacterial Rho helicase: a new tool to dissect mRNP biogenesis and quality control in yeast.* Trends in Cell & Molecular Biology 2014. **9**: p. 79-93.

216. Mosrin-Huaman, C., N. Hervouet-Coste, and A.R. Rahmouni, *Co-transcriptional degradation by the 5'-3' exonuclease Rat1p mediates quality control of HXK1 mRNP biogenesis in S. cerevisiae.* RNA Biology, 2016. **13**(6): p. 582-592.

217. Mosrin-Huaman, C., R. Honorine, and A.R. Rahmouni, *Expression of bacterial Rho factor in yeast identifies new factors involved in the functional interplay between transcription and mRNP biogenesis.* Mol Cell Biol, 2009. **29**(15): p. 4033-4044.

218. Moreau, K., et al., *Recombinant yeast and human cells as screening tools to search for antibacterial agents targeting the transcription termination factor Rho.* The Journal of Antibiotics, 2018.

219. Lee, C.-Y., et al., *Common applications of next-generation sequencing technologies in genomic research.* Translational Cancer Research, 2013. **2**(1): p. 33-45.

220. Meaburn, E. and R. Schulz, *Next generation sequencing in epigenetics: insights and challenges.* Semin Cell Dev Biol, 2012. **23**(2): p. 192-9.

221. Majumder, M. and V. Palanisamy, *Compendium of Methods to Uncover RNA-Protein Interactions In Vivo.* Methods Protoc, 2021. **4**(1).

222. Poulsen, L.D., et al., *SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data.* Rna, 2015. **21**(5): p. 1042-52.

223. Park, P.J., *ChIP–seq: advantages and challenges of a maturing technology.* Nature Reviews Genetics, 2009. **10**(10): p. 669-680.

224. Mardis, E.R., *Next-generation DNA sequencing methods.* Annu Rev Genomics Hum Genet, 2008. **9**: p. 387-402.

225. Bansal, V. and C. Boucher, *Sequencing Technologies and Analyses: Where Have We Been and Where Are We Going?* iScience, 2019. **18**: p. 37-41.

226. Chadeau-Hyam, M., et al., *Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers.* Environmental and Molecular Mutagenesis, 2013. **54**(7): p. 542-557.

227. Eddy, S.R., *What is a hidden Markov model?* Nature biotechnology, 2004. **22**(10): p. 1315-1316.

228. Alharbi, W.S. and M. Rashid, *A review of deep learning applications in human genomics using next-generation sequencing data.* Human Genomics, 2022. **16**(1): p. 26.

229. Strasser, K., et al., *TREX is a conserved complex coupling transcription with messenger RNA export.* Nature, 2002. **417**(6886): p. 304-8.

230. Zenklusen, D., et al., *Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p.* Mol Cell Biol, 2002. **22**(23): p. 8241-53.

231. Jensen, T.H., et al., *The DECD box putative ATPase Sub2p is an early mRNA export factor.* Curr Biol, 2001. **11**(21): p. 1711-5.

232. Nielsen, J., *Yeast Systems Biology: Model Organism and Cell Factory.* Biotechnol J, 2019. **14**(9): p. e1800421.

233. Phillips, S. and J.S. Butler, *Contribution of domain structure to the RNA 3' end processing and degradation functions of the nuclear exosome subunit Rrp6p.* RNA, 2003. **9**(9): p. 1098-107.

234. Hainer, S.J., et al., *Intergenic transcription causes repression by directing nucleosome assembly.* Genes & Development, 2011. **25**(1): p. 29-40.

235. Yu, Y., et al., *Disruption of promoter memory by synthesis of a long noncoding RNA.* Proceedings of the National Academy of Sciences, 2016. **113**(34): p. 9575-9580.

236. van Werven, Folkert J., et al., *Transcription of Two Long Noncoding RNAs Mediates Mating-Type Control of Gametogenesis in Budding Yeast.* Cell, 2012. **150**(6): p. 1170-1181.

237. Nevers, A., et al., *Antisense transcriptional interference mediates condition-specific gene repression in budding yeast.* Nucleic Acids Research, 2018: p. gky342-gky342.

238. Belair, C., et al., *The RNA exosome nuclease complex regulates human embryonic stem cell differentiation.* Journal of Cell Biology, 2019. **218**(8): p. 2564-2582.

239. Kawabe, Y., et al., *The RNA exosome complex degrades expanded hexanucleotide repeat RNA in C9orf72 FTLD/ALS.* The EMBO Journal, 2020. **39**(19): p. e102700.

240. Jain, A. and R.D. Vale, *RNA phase transitions in repeat expansion disorders.* Nature, 2017. **546**(7657): p. 243-247.

241. Lubas, M., et al., *The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis.* Cell Reports, 2015. **10**(2): p. 178-192.

242. Wegener, M. and M. Müller-McNicoll, *Nuclear retention of mRNAs - quality control, gene regulation and human disease.* Semin Cell Dev Biol, 2018. **79**: p. 131-142.

243. Thomsen, R., et al., *Localization of nuclear retained mRNAs in Saccharomyces cerevisiae.* Rna, 2003. **9**(9): p. 1049-57.

244. Paul, B. and B. Montpetit, *Altered RNA processing and export lead to retention of mRNAs near transcription sites and nuclear pore complexes or within the nucleolus.* Molecular Biology of the Cell, 2016. **27**(17): p. 2742-2756.

245. Nalavade, R., et al., *Mechanisms of RNA-induced toxicity in CAG repeat disorders.* Cell Death Dis, 2013. **4**(8): p. e752.

246. Liao, S.E. and O. Regev, *Splicing at the phase-separated nuclear speckle interface: a model.* Nucleic Acids Research, 2020. **49**(2): p. 636-645.

247. Wang, K., et al., *Intronless mRNAs transit through nuclear speckles to gain export competence.* J Cell Biol, 2018. **217**(11): p. 3912-3929.

248. Mure, F., et al., *The splicing factor SRSF3 is functionally connected to the nuclear RNA exosome for intronless mRNA decay.* Sci Rep, 2018. **8**(1): p. 12901.

249. Galganski, L., M.O. Urbanek, and W.J. Krzyzosiak, *Nuclear speckles: molecular organization, biological function and role in disease.* Nucleic Acids Research, 2017. **45**(18): p. 10350-10368.

250. Ilık, İ.A. and T. Aktaş, *Nuclear speckles: dynamic hubs of gene expression regulation.* The FEBS Journal, 2022. **289**(22): p. 7234-7245.

251. Babour, A., et al., *The Chromatin Remodeler ISW1 Is a Quality Control Factor that Surveys Nuclear mRNP Biogenesis.* Cell, 2016. **167**(5): p. 1201-1214.e15.

252. Martinez-Lumbreras, S., et al., *Gbp2 interacts with THO/TREX through a novel type of RRM domain.* Nucleic Acids Res, 2016. **44**(1): p. 437-48.

253. Villa, T., et al., *Degradation of Non-coding RNAs Promotes Recycling of Termination Factors at Sites of Transcription.* Cell Rep, 2020. **32**(3): p. 107942.

254. McWhite, C.D., et al., *Co-fractionation/mass spectrometry to identify protein complexes.* STAR Protoc, 2021. **2**(1): p. 100370.

255. Haglin, E.R., et al., *His-Tag-Mediated Dimerization of Chemoreceptors Leads to Assembly of Functional Nanoarrays.* Biochemistry, 2017. **56**(44): p. 5874-5885.

256. Novačić, A., et al., *Antisense non-coding transcription represses the PHO5 model gene at the level of promoter chromatin structure.* PLoS Genet, 2022. **18**(10): p. e1010432.

257. Fasken, M.B. and A.H. Corbett, *Mechanisms of nuclear mRNA quality control.* RNA Biol, 2009. **6**(3): p. 237-41.

258. Kilchert, C., S. Wittmann, and L. Vasiljeva, *The regulation and functions of the nuclear RNA exosome complex.* Nat Rev Mol Cell Biol, 2016. **17**(4): p. 227-239.

259. Kuai, L., B. Das, and F. Sherman, *A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae.* Proceedings of the National Academy of Sciences, 2005. **102**(39): p. 13962-13967.

260. Ren, Y., et al., *A global screening identifies chromatin-enriched RNA-binding proteins and the transcriptional regulatory activity of QKI5 during monocytic differentiation.* Genome Biology, 2021. **22**(1): p. 290.

Valentin BEAUVAIS

Perturbation de la biogénèse des mRNPs par le facteur bactérien Rho : analyse génomique du recrutement du complexe THO et de ses sous-unités

La transcription des ARN messagers (mRNAs) est un processus complexe qui implique une grande diversité d'acteurs à des étapes bien précise. La production d'un mRNA passe également par sa maturation en particule ribonucleoprotéique (mRNP) par l'ajout de protéines qui vont l'empaqueter, le protéger et le rendre compétent à l'export au cytoplasme ou il sera traduit en protéine. Cette étape est surveillée par un système de contrôle qualité qui détecte les rares mRNPs aberrantes et induit leur dégradation par l'une des exonucléases de l'exosome, Rrp6. Le taux d'erreur lors de la maturation des mRNAs est très faible et ne permet pas l'étude du système de QC. L'induction du facteur bactérien Rho provoque la formation des mRNPs aberrantes nécessaire à l'étude tout en évitant de supprimer d'éventuels acteurs de ce système de QC. Cette étude se focalise sur le complexe THO, une pierre angulaire de la maturation des mRNAs car il est recruté pendant la transcription et sert ensuite de plateforme d'interaction pour d'autres protéines d'empaquetage. L'utilisation de méthode d'analyses sur génome entier (ChIP-seq) a révélé l'implication de la protéine Tho2 dans la détection des transcrits aberrants et leur dégradation par Rrp6, indépendamment du complexe THO. De plus, une approche similaire (RNA-seq) sur des levures dépourvues de Rrp6 a mis une fois de plus en évidence l'importance de la dégradation des ncRNAs dans la régulation de l'expression de gènes codants. Pour finir, l'extension de l'utilisation de Rho depuis la levure chez l'humain pourrait révéler à terme l'implication des granules de stress nucléaires (nuclear speckles) dans le contrôle qualité des mRNAs chez l'humain.

Mots clés : biogénèse des mRNPs, techniques HTS, Rho, complexe THO, Rrp6, Nuclear Speckles

mRNPs biogenesis perturbation by the bacterial Rho factor : genomic analysis of the recrutment of the THO complex and its subunits

The transcription of messenger RNA (mRNA) is a complex process implicating a large diversity of players at very specific steps. The production of an mRNA also requires its maturation into a ribonucleoproteic particle (mRNP) through the binding of proteins that will package, protect and allow the mRNP to be exported to the cytoplasm where it will be translated into a protein. This crucial step is monitored by a quality control system (QC) that detects the rare aberrant mRNPs and induces their degradation by one of the exosome-dependent exonuclease, Rrp6. The error rate during mRNA maturation is very low and complicates the study of the QC. The induction of the bacterial factor Rho leads to the formation of aberrant mRNPs while avoiding the deletion of any protein potentially implicated in the QC system. The present study focuses on the THO complex, a keystone for the maturation of mRNA as it binds to chromatin early during transcription and then serves as a recruitment platform for other packaging proteins. The use of genome-wide analysis techniques (ChIP-seq) revealed the implication of Tho2 for the detection of aberrant transcripts and their degradation by Rrp6, independently from the THO complex. Furthermore, a similar approach (RNA-seq) on Rrp6-depleted yeast highlighted once again the importance of ncRNA degradation for the regulation of coding gene expression. Finally, the expansion of Rho utilization from yeast to human could bring to light the implication of nuclear speckles in the quality control of mRNPs in humans.

Keywords : mRNP biogenesis, HTS techniques, Rho, THO complex, Rrp6, Nuclear speckles

Centre de Biophysique Moléculaire

CNRS UPR4301 groupe « biologie de l'ARN et ARN thérapeutiques »

Rue Charles Sadron Orléans Cedex 2